SEMI-EMPIRICAL PROBABILITY DISTRIBUTIONS AND THEIR APPLICATION IN WAVE-STRUCTURE INTERACTION PROBLEMS

A Dissertation

by

AMIR HOSSEIN IZADPARAST

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2010

Major Subject: Ocean Engineering

SEMI-EMPIRICAL PROBABILITY DISTRIBUTIONS AND THEIR APPLICATION IN WAVE-STRUCTURE INTERACTION PROBLEMS

A Dissertation

by

AMIR HOSSEIN IZADPARAST

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	John M. Niedzwecki
Committee Members,	Kenneth F. Reinschmidt
	James M. Kaihatu
	Steven F. DiMarco
Head of Department,	John M. Niedzwecki

December 2010

Major Subject: Ocean Engineering

ABSTRACT

Semi-empirical Probability Distributions and Their Application in Wave-Structure Interaction Problems. (December 2010)

Amir Hossein Izadparast, B.S., Shiraz University;

M.S., University of Tehran

Chair of Advisory Committee: Dr. John Michael Niedzwecki

In this study, the semi-empirical approach is introduced to accurately estimate the probability distribution of complex non-linear random variables in the field of wave-structure interaction. The structural form of the semi-empirical distribution is developed based on a mathematical representation of the process and the model parameters are estimated directly from utilization of the sample data. Here, three probability distributions are developed based on the quadratic transformation of the linear random variable. Assuming that the linear process follows a standard Gaussian distribution, the three-parameter Gaussian-Stokes model is derived for the second-order variables. Similarly, the three-parameter Rayleigh-Stokes model and the four-parameter Weibull-Stokes model are derived for the crests, troughs, and heights of non-linear process assuming that the linear variable has a Rayleigh distribution or a Weibull distribution. The model parameters are empirically estimated with the application of the conventional method of moments and the newer method of L-moments. Furthermore, the application of semi-empirical models in extreme analysis and estimation of extreme statistics is

discussed. As a main part of this research study, the sensitivity of the model statistics to the variability of the model parameters as well as the variability in the samples is evaluated. In addition, the sample size effects on the performance of parameter estimation methods are studied.

Utilizing illustrative examples, the application of semi-empirical probability distributions in the estimation of probability distribution of non-linear random variables is studied. The examples focused on the probability distribution of: wave elevations and wave crests of ocean waves and waves in the area close to an offshore structure, wave run-up over the vertical columns of an offshore structure, and ocean wave power resources. In each example, the performance of the semi-empirical model is compared with appropriate theoretical and empirical distribution models. It is observed that the semi-empirical models are successful in capturing the probability distribution of complex non-linear variables. The semi-empirical models are more flexible than the theoretical models in capturing the probability distribution of data and the models are generally more robust than the commonly used empirical models.

To my beloved family and friends

ACKNOWLEDGMENTS

I would like to gratefully acknowledge the partial financial support of the Texas Engineering Experiment Station and the R.P. Gregory ’32 Chair endowment during my Ph.D. studies. In addition, I would like to recognize Dr. Per S. Teigen of StatoilHydro and the Offshore Technology Research Center for permission to utilize the experimental data presented in this study.

I would like to express my sincere gratitude and thankfulness to my academic advisor, Prof. John Niedzwecki, for his encouragement, supervision, and support. His great vision and guidance were crucial in the completion of this dissertation. I am also grateful to my advisory committee members: Dr. Kenneth Reinschmidt, Dr. James Kaihatu, and Dr. Steven DiMarco for their input and guidance throughout this research study.

I am deeply thankful to my parents for their immense support, unconditional love, and inspiration throughout my life. I would also like to thank my sister for her constant encouragement. Finally, I am greatly thankful to my Noushin for her patience, support, and help. She has given me emotional solace and has been my greatest inspiration.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

LIST OF FIGURES

Page

Page

Page

Page

Page

LIST OF TABLES

Page

Page

Page

CHAPTER I

INTRODUCTION

Ocean engineers are routinely involved in a wide range of design problems that involve in both extreme and normal environmental conditions. One of the most challenging problems in the field of ocean engineering is the interaction of ocean waves with offshore and coastal structures. The interaction between highly energetic waves and the structure may result in unanticipated excessive structural motions or loadings resulting from green water and wave deck impact. At some offshore sites, mildly energetic ocean waves may be converted to usable energy, utilizing Wave Energy Convertor (WEC) devices. The random characteristic of ocean environment requires engineers to consider the effects of random variability of the pertinent variables in their predictive models and design processes. Thus for offshore design, one needs to have accurate estimates of the probability distribution of the key random variables. The probability distributions may be used to estimate the reliability of an engineering design or the risk associated with an expectation.

In the theory of statistics, parametric distribution models are considered as a major family of probability distributions in which a distribution function consists of an underlying structural form that is dependent upon finite number of parameters. There are

---

This dissertation follows the style of Journal of Applied Ocean Research.

two commonly used approaches to specify the structural form and estimate the parameters of a parametric model:

1) Theoretical Approach: Here, the form and parameters of the model are derived based on a mathematical model that approximates the physics governing the process. Examples of some studies on the theoretical probability distributions of random variables in the field of ocean engineering are presented in Table 1. The main advantage of the theoretical approach is that the probability distribution typically reflects some physical insight and the parameter estimation requires only limited information about the process.

2) Empirical Approach: Here, it is assumed that the random variable follows a standard probability distribution e.g. Gaussian, Weibull, Rayleigh, etc. and the unknown parameters are estimated empirically using sample data. Examples of studies on application of empirical distributions in ocean engineering problems are presented in Table 2. The empirical models have higher flexibility in capturing the probability distribution of data and are usually more efficient as compared to the theoretical distributions. However, the distribution structure and the model parameters of an empirical probability distribution in most cases do not have a clear theoretical interpretation.

In earlier studies, the probability distribution of random variables was commonly estimated from application of theoretical models. Due to availability of data from advanced calibrated numerical models, experimental studies, and full-scale measurements, the empirical models are considered as the more robust options in the

more recent studies. Here, a new approach is introduced in which the theoretical and empirical approaches are combined to more accurately model the probability distribution of random variables in the field of wave-structure interaction. The application of semi-empirical models is studied for random variables in both normal and extreme environmental conditions.

Table 1 Examples of studies performed on theoretical probability distributions.

Random Variable	Reference
Linear Wave heights	[1-5]
Linear wave crests and troughs	[6]
Nonlinear surface wave elevation	[7-10]
Nonlinear wave heights	[11]
Nonlinear wave crests	[8, 12-16]
Nonlinear wave run-up	[17]

Table 2 Examples of studies performed on empirical probability distributions.

Random Variable	Reference
Nonlinear Wave Crests and Heights	[18-21]
Nonlinear Run-up	[22, 23]
Wave power	[24, 25]

I.1 Problem Definition

A wave-structure problem can be considered as a stochastic input/output system in which the input is the incident wave and the outputs are the responses, e.g. wave run-up, wave forces, structural motions, etc. In general, both system input and output may be non-linear processes. The non-linearity may be due to the nature of the relationship of the random variables, e.g. the quadratic relation between wave height and wave power, or may be caused by the interaction of linear process with higher order terms, e.g. non-linear interaction of surface waves. For engineering purposes, linear models are commonly used to approximate the non-linear variables and consequently simplify the calculations. However, the linear approximations cannot adequately model the non-linear behavior of random variables in extreme environments or when the process is non-linear in nature.

In order to accurately model the non-linear processes, complex mathematical models are required which makes it nearly impossible to derive an exact theoretical probability distribution. Therefore, most theoretical probability distributions are developed based on simplifying assumptions and the models are only applicable for specific conditions. Engineers routinely utilize the empirical models estimate the probability distribution of the non-linear random variables utilizing the sample data. These models are used to predict the probability distribution outside the sample range. The empirical probability distributions have been found fairly accurate in representing the probability distribution of non-linear processes. However, the selection of the distribution structure is an iterative process and in some cases is difficult to justify. This

is mainly because the connection between the structural form of the empirical probability distribution and the physics underlying the process is not clear.

In this study a new approach, i.e. semi-empirical approach, is introduced to model the probability distribution of complex random variables in the field of wave-structure interaction. An important aspect of this research is that the structural form of the semi-empirical probability distribution is derived, similar to the theoretical approach, utilizing a mathematical model and application of random variable theory. The model parameters are empirically estimated utilizing sample data sets. Therefore, the semi-empirical probability distribution takes advantage of both the structural form and incorporation of parameters based upon physical interpretation of the process. The utilization of empirically estimated parameters improves the flexibility of the model to capture the actual probability distribution of data.

I.2 Research Methodology

This research study addresses three major aspects: 1) the development of the semi-empirical model, 2) the analysis of model uncertainty, and 3) the application of the model and the evaluation of its performance in comparison to other models.

I.2.1 Model development

The first step in developing a semi-empirical model is deriving the structural form of the distribution. Similar to a theoretical probability distribution, the structural form of a semi-empirical probability distribution is developed from a mathematical model that approximates the random process of interest. The mathematical model

essentially correlates the unknown random variable to the random variables with known probability distributions. The random variable transformation theory [26] is then applied to derive the probability distribution of the new random variable. Assuming that the mathematical model is a reasonable approximation of the actual process, one could reasonably expect that the semi-empirical probability model is a robust representative of the variability of the random process.

In this research study, a second-order perturbation expansion is used to relate the non-linear random variable to the corresponding linear process. Second-order approximations actually cover a wide range of weakly non-linear random variables in the field of wave-structure interaction. The mathematical model can be considered as a quadratic transformation of the linear process. The three probability distributions, i.e. Gaussian-Stokes (GS), Rayleigh-Stokes (RS), and Weibull-Stokes (WS), are developed assuming that the first-order random variable has a Gaussian, Rayleigh, or Weibull probability distribution, respectively. These later distributions have been widely used in the field of ocean engineering to model the probability distribution of linear random variables. It is well-known that the observations of a linear process have a Gaussian probability distribution [27]. Longuet-Higgins [1] showed that the crests and troughs, maximum and minimum amplitudes in between two consecutive zero-upcrossing, of a narrow-banded linear random process follow the Rayleigh distribution law. The Weibull distribution has also been used to model the probability distribution of crests and troughs of linear processes with finite spectral width (see e.g. [18]). In the timeseries analysis of random processes, four random variables are commonly considered: individual

observations of surface wave elevations η, the wave crests a^+, the wave troughs a^-, or the wave heights H, each of identified in Figure 1. The GS model is appropriate for modeling the probability distribution of wave elevations of a weakly non-linear process and the RS and WS models can be used to estimate the probability distributions of wave crests, wave troughs, and wave heights. Both GS and RS models are three-parameter distributions while the WS is a more general form of the RS model and has an additional shape parameter.

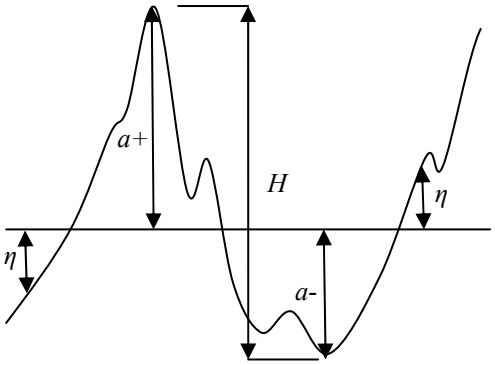


Figure 1 A schematic representation of the random variables in a timeseries.

The next step in the semi-empirical model development is the empirical parameter estimation in which the parameters are related directly to the sample statistics. In the literature, there are numerous parameter estimation methods; examples included the method of maximum likelihood, the method of least squares, the method of moments, the method of L-moments, and the Bayesian model approach. An introduction to the parameter estimation methods can be found in [28, 29]. Among the parameter estimation methods, the method of moments (MoM) is probably the most popular in the

field of ocean engineering. In this method the model parameters are obtained from equating the distribution moments with their corresponding sample statistics. A relatively newer parameter estimation method is the method of L-moments (MoL) which was developed by Hosking [30]. It is similar to method of moments in that the model parameters are estimated from equating the distribution statistics with their corresponding sample estimates which in this case are the L-moments. Distribution L-moments are developed from probability weighted moments formerly introduced by Greenwood et al. [31]. The sample L-moments are obtained from a specific linear combination of data points which make them less influenced by extraordinary large sample points, as compared to the ordinary moments. The method of L-moments has been applied to different probability distributions and its performance has been compared with other parameter estimation methods [29, 32, 33, 34-36]. The results of these studies indicate that the MoL is a robust and efficient parameter estimation method, especially for extreme analysis and for samples with limited size. The MoL has been used for extreme analysis of data sets from different engineering fields [37, 38]. In this study both MoM and MoL are applied to obtain the relation between the distribution statistics and model parameters and eventually to estimate the model parameters.

I.2.2 Model uncertainty analysis

It is crucial to identify the sources of uncertainty and to evaluate the effects of uncertainties on the model results. A major source of uncertainty in the estimates of a semi-empirical model is the mathematical model being utilized to develop the distribution structure. A challenging task is to make sure that the mathematical model is

a reasonable approximation of the actual process and this requires a fairly comprehensive knowledge of the process. In the derivation of the probability distribution, it is assumed that the model parameters are deterministic constants; however, these parameters contain uncertainty and their variability affects the estimates of the semi-empirical probability distribution. In this research study, these effects are evaluated utilizing illustrative numerical examples.

Another major source of uncertainty of the semi-empirical models comes from the empirical parameter estimation. Sample variability, error in the parameter estimation, and uncertainty of the measurements are the main causes of uncertainty of the empirical parameter estimation. A sample data set represents a part of the entire domain and consequently the variability of the sample's statistics is inevitable which directly causes variability in the empirically estimated model parameters. In this study, the effects of sample variability and empirical parameter estimation uncertainty on the estimates of the semi-empirical models are evaluated utilizing the Monte-Carlo simulation technique. Additionally, the sample size effects on the variability of the model statistics are studied.

There are known and unknown errors associated with each parameter estimation method which causes uncertainty in the empirical parameter estimation. The efficiency of the parameter estimation methods, i.e. MoL and MoM for different conditions are compared utilizing the Monte-Carlo samples. Additionally, the measurement quality is always a major concern in any empirical model estimation. In this study numerical examples are utilized to show the effects of the faulty measurements on the estimates of the semi-empirical models.

I.2.3 Model application and evaluation

In the final chapter, the semi-empirical probability distributions are utilized to estimate the probability distribution of non-linear random variables involved in the wave-structure interaction problem. As examples, the GS model is applied to estimate the probability distribution of wave elevations of undisturbed ocean waves and disturbed waves interacting with an offshore structure. The RS and WS models are used to model the probability distribution of wave crests far away from any obstacle and wave crests in the vicinity of and beneath an offshore structure. In another case, the RS and WS models are utilized to estimate the probability distribution of wave run-up over the vertical columns of the offshore platform. A special form of the non-linear transformation with no linear term is applied for wave power in benign environments. In each example, the statistics of the semi-empirical model are compared with appropriate theoretical and empirical probability distributions. The sample data utilized in this research investigation were mainly obtained from experimental studies on behavior of offshore platforms, and when necessary, samples were generated numerically.

The performance of the semi-empirical model in capturing the probability distribution of sample data is evaluated utilizing bootstrap analysis. Bootstrap analysis is a powerful simulation technique typically used to determine the bias and variance associated with a sample estimate of a parameter of interest [39-40]. In this method independent samples of the same size as the original sample are generated from a distribution model with replacement. In the semi-parametric bootstrap the distribution model is approximated by a smoothed probability distribution utilizing peaks-over-

threshold (POT) method for tail modification [41]. The smoothing improves the convergence of bootstrap estimates and the tail modification enables generation of sample points beyond the original sample range. Here, the root-mean-square errors (RMSE) of the model estimates obtained by bootstrap re-sampling method are utilized as metrics to quantitatively evaluate the model performance.

CHAPTER II

SEMI-EMPIRICAL MODEL DEVELOPMENT

II.1 Mathematical Background

Based on perturbation theory, an approximate solution for a weakly non-linear process ζ_n can be obtained by the expansion,

$$\zeta_n = \varepsilon^0 \zeta^{(0)} + \varepsilon^1 \zeta^{(1)} + \varepsilon^2 \zeta^{(2)} + \cdots \tag{2.1}$$

where ε is a small term ($\varepsilon \ll 1.0$), and $\zeta^{(m)}$ is the term of order m. The approximation ensures that the contribution of the term $\varepsilon^m \zeta^{(m)}$ is significantly less than its previous terms, i.e. $\varepsilon^1 \zeta^{(1)}, \ldots, \varepsilon^{m-1} \zeta^{(m-1)}$. For specific types of boundary value problem, e.g. the water wave boundary value problem [42], the approximation can be rewritten as a power series in terms of the linear process ζ, specifically,

$$\zeta_n = \delta_0 \zeta^0 + \delta_1 \zeta^1 + \delta_2 \zeta^2 + \cdots \tag{2.2}$$

where δ_m is the amplification of the term of order m. To satisfy the basic assumption of the perturbation expansion the amplification factors in Eq. (2.2) should satisfy $\delta_{m+1} \ll \delta_m$ for any $m > 0$. The second-order transformation is defined as,

$$\zeta_n = \gamma + \alpha \zeta + \beta \zeta^2 \tag{2.3}$$

where γ indicates the constant linear shift between the linear and non-linear process, i.e. zeroth-order, α and β specify the amplification of the first- and second-order terms,

respectively. In this research study, the second-order approximation is utilized which covers a wide range of problems in wave-structure interaction.

Knowing the probability density function (PDF) of the linear process f_ζ, one can obtain the PDF of the non-linear process from the random variables transformation rule as [26],

$$f_{\zeta_n}(x) = f_\zeta\left(\zeta = G(\zeta_n), \zeta_n = x\right)\left|\frac{\partial}{\partial \zeta_n}\left(G(\zeta_n), \zeta_n = x\right)\right| \tag{2.4}$$

where the transformation function $G(\zeta_n)$ is given as,

$$G(\zeta_n) = \frac{-\alpha \pm \left(\alpha^2 + 4\beta(\zeta_n - \gamma)\right)^{1/2}}{2\beta} \tag{2.5}$$

Note that in the second-order transformation, it is assumed that $\alpha \geq 0$, β and γ are real numbers, and $|\beta| \ll \alpha$.

A special case is the model with no linear term, i.e. $\alpha = 0$ where the transformation function $G(\zeta_n)$ simplifies into,

$$G(\zeta_n) = \pm\left(\frac{(\zeta_n - \gamma)}{\beta}\right)^{1/2} \tag{2.6}$$

where without loss of generality, it is assumed that $\beta > 0$. Based on the probability distribution of the linear random variable ζ, different distribution models are developed for ζ_n. Knowing the PDF of f_{ζ_n}, the cumulative distribution function (CDF) of ζ_n is obtained from the definition,

$$F_{\zeta_n}(x) = \int_0^x f_{\zeta_n}(x')\,dx' \tag{2.7}$$

The quantile function, i.e. the inverse form of CDF, is another useful form of the probability distribution. For any given probability of $0 < u < 1$, the quantile function $x_{\zeta_n}(u)$ is the unique value that satisfies,

$$F_{\zeta_n}\left(x_{\zeta_n}(u)\right) = u \tag{2.8}$$

It is common to define the quantile function in terms of the probability of exceedance $P = 1 - u$.

II.2 Empirical Parameter Estimation

II.2.1 Method of moments

Distribution moments have been widely used to characterize the probability distributions and the method of moments has been commonly applied as an efficient parameter estimation technique. For a random variable X, the first moment, i.e. mean, is defined as [32],

$$\mu_1(X) = E(X) \tag{2.9}$$

and the nth moment is defined as,

$$\mu_n(X) = E(X - \mu)^n \tag{2.10}$$

where, $E\left(g(X)\right)$ is the expectation of the function $g(X)$ and is defined as,

$$E\left(g(X)\right) = \int_{-\infty}^{\infty} g(x) f(x)\,dx = \int_{-\infty}^{\infty} g(x)\,dF(x) = \int_0^1 g\left(x(u)\right)du \tag{2.11}$$

From that, the distribution moments can be defined in the form of,

$$\begin{aligned} \mu_1(X) &= \int_0^1 x(u)\,du & n=1 \\ \mu_n(X) &= \int_0^1 \left(x(u)-\mu_1\right)^n du & n>1 \end{aligned} \tag{2.12}$$

The mean represents the centroid of the distribution and the variance $\sigma_X{}^2 = \mu_2(X)$ is a measure of the distribution dispersion around its center. Other useful moments are the dimensionless third and fourth moments, respectively called skewness $s_X = \mu_3(X)/\sigma_X{}^3$ and excess kurtosis $K_X = \mu_4(X)/\sigma_X{}^4 - 3$.

In order to estimate the m parameters of a probability distribution with method of moments, the first m distribution moments are equated with their unbiased sample estimates, which give m equations to be solved for the m unknown parameters. The sample estimates of the distribution moments can be obtained from a data set $x_1, x_2, \ldots, x_N$ of size N. Particularly, the sample mean is estimated as,

$$\hat{\mu}_1(X) = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2.13}$$

and the sample estimate of the nth moment $n>1$ can be obtained from,

$$\hat{\mu}_n(X) = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \hat{\mu}_1\right)^n \tag{2.14}$$

The sample mean, defined in Eq. (2.13), is an unbiased estimator of the distribution mean while the higher order sample moments are biased estimators. The unbiased sample estimators of the second, third, and fourth moments are estimated, respectively from,

$$\hat{\mu}_2(X) = \hat{\sigma}_X{}^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}\left(x_i - \hat{\mu}_1\right)^2 \tag{2.15}$$

$$\hat{\mu}_3(X) = \frac{N}{(N-1)(N-2)} \sum_{i=1}^{N} (x_i - \hat{\mu}_1)^3 \tag{2.16}$$

$$\hat{\mu}_4(X) = \frac{N(N+1)}{(N-1)(N-2)(N-3)} \sum_{i=1}^{N} (x_i - \hat{\mu}_1)^4 \tag{2.17}$$

II.2.2 Method of L-moments

The application of L-moments for parameter estimation purposes was initiated by Hosking [30]. L-moments are developed as modifications of the probability-weighted moments (PWM) formerly introduced by Greenwood et al. [31]. For a random variable X with quantile function of $x(u)$, two special forms of the nth PWM are defined as,

$$\mu_{PW,1,n}(X) = \int_0^1 x(u)(1-u)^n \, du, \qquad \mu_{PW,2,n}(X) = \int_0^1 x(u) u^n du \tag{2.18}$$

As shown here, the PWMs are developed from integration of successively higher powers of u (or $1-u$), whereas ordinary moments, (2.12) are developed from integration of successively higher powers of the quantile function. The main difference between moments and PWMs is that moments give greater weight to the extreme tails of the distribution, since for most distributions $x(u)$ increases much faster than u as u approaches unity. This helps the sample PWMs to be less affected by unexpectedly large (or small) values and consequently the high order sample PWMs are less biased as compared with the corresponding ordinary moments.

PWMs, as alternatives to ordinary moments, have been used in the field of probability distribution parameter estimation e.g. [43, 44]. However, it is difficult to directly connect PWMs to the characteristics of the probability distribution, e.g. shape

and scale. Hosking [30] introduced L-moments from a linear combination of PWMs to overcome this issue. For a random variable X with quantile function of $x(u)$, the distribution L-moments are obtained from,

$$\lambda_n(X) = \int_0^1 x(u) P^*_{n-1}(u)\, du \tag{2.19}$$

which resembles the PWM, defined in Eq. (2.18), in which the polynomial $(1-u)^n$ is replaced by the shifted Legendre polynomials $P^*_{n-1}(u)$,

$$P^*_n(u) = \sum_{k=0}^{r} p^*_{n,k}\, u^k \tag{2.20}$$

where,

$$p^*_{n,k} = \frac{(-1)^{n-k}(n+k)!}{(k!)^2 (n-k)!} \tag{2.21}$$

In terms of the PWMs $\mu_{PW,1,n}$ the first four L-moments are given as,

$$\begin{aligned}
\lambda_1(X) &= \mu_{PW,1,0} \\
\lambda_2(X) &= \mu_{PW,1,0} - 2\mu_{PW,1,1} \\
\lambda_3(X) &= \mu_{PW,1,0} - 6\mu_{PW,1,1} + 6\mu_{PW,1,2} \\
\lambda_4(X) &= \mu_{PW,1,0} - 12\mu_{PW,1,1} + 30\mu_{PW,1,2} - 20\mu_{PW,1,3}
\end{aligned} \tag{2.22}$$

The dimensionless forms of the distribution L-moments are useful statistics which are defined as,

$$\tau(X) = \frac{\lambda_n(X)}{\lambda_1(X)} \qquad n = 2, \qquad \tau_r = \frac{\lambda_n(X)}{\lambda_2(X)} \qquad n > 2 \tag{2.23}$$

By definition, λ_1 is the L-location or mean of the distribution, λ_2, τ, τ_3, and τ_4 are L-scale, L-CV, L-skewness, L-kurtosis, respectively, and are analogous to the ordinary standard deviation, coefficient of variation, skewness, and kurtosis. A more complete definition of the L-moments and their characteristics can be found in [30].

Similar to the method of moments, in order to estimate the m parameters of a probability distribution with method of L-moments, the first m distribution L-moments are equated with their unbiased sample estimates. The sample L-moment l_n of an ordered sample $x_{1:N} \le x_{2:N} \le \ldots \le x_{N:N}$ of size N is defined as,

$$l_{n+1}(X) = \sum_{k=0}^{n} p^*_{n,k} \hat{\mu}_{PW,2,k}(X) \qquad n = 0, 1, ..., N-1 \tag{2.24}$$

where $\hat{\mu}_{PW,2,n}(X)$ is an unbiased estimator of $\mu_{PW,2,n}(X)$, defined in Eq. (2.18). Specifically,

$$\hat{\mu}_{PW,2,n}(X) = N^{-1} \binom{N-1}{n}^{-1} \sum_{j=n+1}^{N} \binom{j-1}{n} x_{j:N} \tag{2.25}$$

and the notation in brackets denotes binomial coefficients, for example

$$\binom{k}{j} = \frac{k!}{j!\,(k-j)!} \qquad k \ge j \tag{2.26}$$

Note that, the sample L-moment l_n is an unbiased estimator of the distribution L-moments λ_n.

II.3 Gaussian-Stokes Model

II.3.1 Model development

It is well known that the linear random process has a standard Gaussian probability distribution [27], specifically,

$$f_{\zeta}(x)=\frac{1}{2\pi}\exp\left(-\frac{x^2}{2}\right) \qquad -\infty<x<\infty \tag{2.27}$$

Utilizing the second-order transformation defined in Eq. (2.3), it can be shown that the PDF of the non-linear process ζ_n in the Gaussian-Stokes (GS) model is in the form of,

$$\begin{aligned} f_{\zeta_n}(x)&=\frac{1}{\sqrt{2\pi}}\chi^{-1}\left[\exp\left(-\frac{(\chi-\alpha)^2}{8\beta^2}\right)+\exp\left(-\frac{(\alpha+\chi)^2}{8\beta^2}\right)\right] \\ \gamma-\frac{\alpha^2}{4\beta}<x<\infty, &\qquad \text{for } \beta>0 \\ -\infty<x<\gamma-\frac{\alpha^2}{4\beta} &\qquad \text{for } \beta<0 \end{aligned} \tag{2.28}$$

where,

$$\chi=\left(\alpha^2+4\beta(x-\gamma)\right)^{1/2} \tag{2.29}$$

and $f_{\zeta_n}(x)=0$ for x outside the defined range. Note that $\zeta_n=\gamma-\frac{\alpha^2}{4\beta}$ is equivalent to $\zeta=\frac{-\alpha}{2\beta}$ and since it is assumed that $|\beta|\ll\alpha$, the $\text{Prob}\left(x\le\frac{-\alpha}{2\beta}\right)$ for $\beta>0$ and $\text{Prob}\left(x\ge\frac{-\alpha}{2\beta}\right)$ for $\beta<0$ are negligible. There is no closed form solution for the CDF of

the ζ_n in the GS model. However, the distribution may be presented in terms of the Gaussian CDF, $\Phi(x)=\frac{1}{2\pi}\int_{-\infty}^{x}\exp\left(-y^2/2\right)dy$, as,

$$F_{\zeta_n}(x)=\Phi\left(\frac{\chi-\alpha}{2\beta}\right)+\Phi\left(\frac{\chi+\alpha}{2\beta}\right) \tag{2.30}$$

An approximate estimate for the quantile function can be obtained knowing that $(\alpha-\chi)\ll(\chi+\alpha)$ and consequently the second term in the $F_{\zeta_n}(x)$ can be ignored with respect to the first term. Therefore,

$$x_{\zeta_n}(u)\approx\gamma+\beta\left(\Phi^{-1}(u)\right)^2+\alpha\ \Phi^{-1}(u) \tag{2.31}$$

where, $\Phi^{-1}(u)$ is the inverse function (quantile function) of $\Phi(x)$.

For the special case with no linear term, i.e. $\alpha=0$, the PDF of ζ_n in the GS model is derived as,

$$f_{\zeta_n}(x)=\frac{1}{\left(2\pi\beta(x-\gamma)\right)^{1/2}}\exp\left(-\frac{(x-\gamma)}{2\beta}\right),\qquad \frac{x-\gamma}{\beta}>0 \tag{2.32}$$

II.3.2 Parameter estimation

II.3.2.1 Method of moments

The distribution moments of the linear process ζ with Gaussian probability distribution are,

$$\begin{aligned}&\mu_r(X)=0 && \text{n is odd}\\ &\mu_r(X)=(n-1)!! && \text{n is even}\end{aligned} \tag{2.33}$$

where the !! indicates double factorial. From this, the first four distribution moments of ζ are obtained as,

$$\mu_1(\zeta)=0,\quad \mu_2(\zeta)=1,\quad \mu_3(\zeta)=0,\quad \mu_4(\zeta)=3 \tag{2.34}$$

Based on these, the first three moments of ζ_n modeled by GS distribution are estimated as,

$$\begin{aligned}
&\mu_1(\zeta_n)=E(\zeta_n)=\gamma+\beta,\\
&\mu_2(\zeta_n)=E\left[\left(\zeta_n-\mu_1(\zeta_n)\right)^2\right]=\alpha^2+2\beta^2,\\
&\mu_3(\zeta_n)=E\left[\left(\zeta_n-\mu_1(\zeta_n)\right)^3\right]=6\alpha^2\beta+8\beta^3.
\end{aligned} \tag{2.35}$$

Substituting the first three sample moments defined respectively in Eq.s (2.13), (2.15), and (2.16) in these equations, empirical estimates of the three parameters $\hat{\alpha},\hat{\beta},$ and $\hat{\gamma}$ can be obtained from simultaneous solution of the system of equations,

$$\begin{aligned}
&4\hat{\beta}^3-6\hat{\beta}\hat{\mu}_2(\zeta_n)+\hat{\mu}_3(\zeta_n)=0,\\
&\hat{\alpha}=\left(\hat{\mu}_2(\zeta_n)-2\hat{\beta}^2\right)^{1/2},\\
&\hat{\gamma}=\hat{\mu}_1(\zeta_n)-\hat{\beta}.
\end{aligned} \tag{2.36}$$

In order to obtain the estimates of the model parameters, firstly, the first relation in Eq. (2.36) needs to be solved for $\hat{\beta}$ and then the value of $\hat{\beta}$ will be used in the second and third relations to estimate $\hat{\alpha}$ and $\hat{\gamma}$. Assuming that $\beta\approx O(\varepsilon)$ where ε is an arbitrary small value, an approximate solution to Eq. (2.36) accurate to $O\left(\varepsilon^2\right)$ is obtained as,

$$\hat{\beta}=\frac{\hat{\mu}_3(\zeta_n)}{6\hat{\mu}_2(\zeta_n)} \tag{2.37}$$

The same relations given in Eq. (2.35) can be used for the special GS model with $\alpha = 0$ and the empirical estimates of the two unknown parameters are obtained in the form of,

$$\begin{aligned} &\alpha = 0, \\ &\hat{\beta} = \left(\frac{\hat{\mu}_2(\zeta_n)}{2} \right)^{1/2}, \\ &\hat{\gamma} = \hat{\mu}_1(\zeta_n) - \hat{\beta}. \end{aligned} \tag{2.38}$$

II.3.2.2 Method of L-moments

Since the closed form of the quantile function $\zeta_n(u)$ of the GS model is not available, the exact analytical formulation of the distribution L-moments cannot be obtained. Therefore, approximate relations between distribution L-moments and the parameters α , β, and γ are obtained applying Monte-Carlo simulations. For this purpose it is assumed that the effect of β on the second L-moments is negligible. Additionally, it is assumed that the linear term does not contribute to the third L-moments. The results of the simulations for different combination of the parameters α , β, and γ with $\alpha/|\beta| > 3$, indicate that,

$$\begin{aligned} &\lambda_1(\zeta_n) = \gamma + \beta, \\ &\lambda_2(\zeta_n) \approx \alpha \, \pi^{-1/2}, \\ &\lambda_3(\zeta_n) \approx 0.5511 \beta. \end{aligned} \tag{2.39}$$

From this set of equations the estimates of the model parameters may be obtained from,

$$\begin{aligned} &\hat{\beta} = l_3(\zeta_n)/0.5511, \\ &\hat{\alpha} = l_2(\zeta_n) \pi^{1/2}, \\ &\hat{\gamma} = l_1(\zeta_n) - \hat{\beta}. \end{aligned} \tag{2.40}$$

For the GS model with no linear term $\alpha = 0$, the relations given in Eq. (2.39) are not applicable. For this case, approximate estimate of L-moments are obtained as,

$$\begin{aligned} &\lambda_1(\zeta_n) = \gamma + \beta, \\ &\lambda_2(\zeta_n) \approx 0.6368\beta. \end{aligned} \tag{2.41}$$

and the estimates of $\hat{\beta}$ and $\hat{\gamma}$ can be easily obtained from equating these two distribution L-moments with the corresponding sample L-moments $l_1(\zeta_n)$ and $l_2(\zeta_n)$.

II.4 Rayleigh-Stokes Model

II.4.1 Model development

The Rayleigh-Stokes (RS) model is basically developed for crests, troughs, and heights of a non-linear random process. It is well known that the crests and troughs of a narrow-band linear random process have a Rayleigh distribution [1], specifically,

$$f_{\zeta}(x) = \frac{x}{R}\exp\left(-\frac{x^2}{2R}\right), \qquad x > 0 \tag{2.42}$$

where R is the Rayleigh parameter. Utilizing the random variable transformation defined in Eq. (2.4), the PDF, CDF, and quantile function of ζ_n for $\beta > 0$ and $x > \gamma$ in the RS model are obtained as,

$$\begin{aligned} f_{\zeta_n}(x) &= \frac{\chi - \alpha}{2\beta R\chi}\exp\left(-\frac{(\chi-\alpha)^2}{8R\beta^2}\right), \qquad x > \gamma \\ F_{\zeta_n}(x) &= 1 - \exp\left(-\frac{(\chi-\alpha)^2}{8R\beta^2}\right), \\ x_{\zeta_n}(u) &= \gamma - 2\beta R\,\ln(1-u) + \alpha\left(-2R\ln(1-u)\right)^{1/2}. \end{aligned} \tag{2.43}$$

where χ is defined in Eq. (2.29). For $\beta < 0$ the distributions become,

$$f_{\zeta_n}(x) = \frac{1}{2\beta R\chi}\left[(\chi-\alpha)\exp\left(-\frac{(\chi-\alpha)^2}{8R\beta^2}\right)H_\gamma(x) + \right.$$
$$\left. -(\chi+\alpha)\exp\left(-\frac{(\chi+\alpha)^2}{8R\beta^2}\right)\right], \tag{2.44}$$
$$F_{\zeta_n}(x) = \left[1-\exp\left(-\frac{(\chi-\alpha)^2}{8R\beta^2}\right)\right]H_\gamma(x) + \exp\left(-\frac{(\chi+\alpha)^2}{8R\beta^2}\right).$$

where $H_\gamma(x)$ is the step function,

$$H_\gamma(x) = \begin{cases} 1 & x \geq \gamma \\ 0 & x < \gamma \end{cases} \tag{2.45}$$

The distributions in Eq. (2.44) are defined for $-\infty < x < \gamma - \frac{\alpha^2}{4\beta}$. The exact analytical form of the quantile function is not available for the RS model with $\beta < 0$. An approximate estimate for the quantile function can be obtained knowing that $(\alpha - \chi) \ll (\chi + \alpha)$ and consequently the third term in the $F_{\zeta_n}(x)$ can be ignored with respect to the second term. As $\chi \to 0$ the terms limit to $\exp\left(-\frac{\alpha^2}{8R\beta^2}\right)$ which is insignificant considering that α^2/β^2 is a significantly large number. Considering these, the approximate estimate of the quantile function is derived as,

$$x_{\zeta_n}(u) \approx \gamma - 2\beta R\,\ln(1-u) + \alpha\left(-2R\ln(1-u)\right)^{1/2} \tag{2.46}$$

For the transformation with no linear term, i.e. $\alpha = 0$, the distribution of ζ_n is the well-known exponential distribution, specifically,

$$
\begin{aligned}
f_{\zeta_n}(x) &= \frac{1}{2\beta R}\exp\left(-\left(\frac{x-\gamma}{2\beta R}\right)\right), \qquad x > \gamma \\
F_{\zeta_n}(x) &= 1-\exp\left(-\left(\frac{x-\gamma}{2\beta R}\right)\right), \\
x_{\zeta_n}(u) &= \gamma - 2\beta R\,\ln(1-u).
\end{aligned}
\tag{2.47}
$$

II.4.2 Parameter estimation

As shown in Eq.s (2.43) and (2.44) the RS distributions have four unknown parameters, $\alpha, \beta, \gamma,$ and R. However, not all the four parameters have independent effects on the distributions and the parameters $\alpha, \beta,$ and R can be merged into two independent parameters. Without loss of generality, it can be assumed that R is known, e.g. $R = 1.0$, and its contribution will be represented by the parameters α and β.

II.4.2.1 Method of moments

The nth raw distribution moment of the Rayleigh distribution can be obtained from,

$$
E\left[\zeta^n\right] = \frac{n}{2}(2R)^{n/2}\,\Gamma\left(\frac{n}{2}\right) \tag{2.48}
$$

Utilizing this and the transformation given in Eq. (2.3), the first three moments of the non-linear process ζ_n in the RS model are obtained in the form of,

$$
\begin{aligned}
\mu_1(\zeta_n) &= \gamma + 2\beta R + \alpha(2R)^{1/2}\,\Gamma(3/2), \\
\mu_2(\zeta_n) &= 4\beta^2R^2 + \alpha\beta(2R)^{3/2}\,\Gamma(3/2) + 2\alpha^2R(1-\pi/4), \\
\mu_3(\zeta_n) &= 16\beta^3R^3 + 9(2)^{1/2}\,\alpha\beta^2\,R^{5/2}\,\Gamma(3/2) + \\
&\qquad 12\alpha^2\beta R^2\left(1-\Gamma^2(3/2)\right) + \alpha^3(2R)^{3/2}\left(2\Gamma^3(3/2) - \frac{3}{2}\Gamma(3/2)\right).
\end{aligned}
\tag{2.49}
$$

As shown in these equations, the nth moment can be defined as a polynomial function of α and β with the same order of n. The empirical estimates of the RS model parameters $\hat{\alpha}, \hat{\beta}, \text{and } \hat{\gamma}$ can be obtained from equating the distribution moments in Eq. (2.49) with their corresponding sample estimates given in Eq.s (2.13), (2.15), and (2.16) and solving the system of equations for the parameters. The procedure requires a non-linear iterative solver and as initial values one may assume that $\hat{\beta}_{initial} = \pm \hat{\alpha}_{initial} / \alpha'$ where α' is an arbitrary large value, e.g. $\alpha' = 10$. From that, the initial values for the parameter α and γ may be obtained from,

$$\begin{aligned} \hat{\alpha}_{initial} &= \left(\frac{\hat{\mu}_2(\zeta_n)}{2R(1-\pi/4) \pm \alpha'^{-1}(2R)^{3/2}\Gamma(3/2)} \right)^{1/2}, \\ \hat{\gamma}_{initial} &= \hat{\mu}_1(\zeta_n) - \hat{\alpha}(2R)^{1/2}\Gamma(3/2) - 2\hat{\beta}R. \end{aligned} \tag{2.50}$$

Note that the relations for $\mu_2(\zeta_n)$ and $\mu_3(\zeta_n)$ given in Eq. (2.49) can be approximated by neglecting the terms with high order of β, i.e. β^2 and β^3, with respect to the other terms. The relations in Eq. (2.49) can be used for the moments of the special model with $\alpha = 0$, Eq. (2.47), and the empirical estimate of the parameters for this model may be obtained from,

$$\begin{aligned} &\alpha = 0, \\ &\hat{\beta} = \frac{\left(\hat{\mu}_2(\zeta_n)\right)^{1/2}}{2R}, \\ &\hat{\gamma} = \hat{\mu}_1(\zeta_n) - 2\hat{\beta}R \end{aligned} \tag{2.51}$$

II.4.2.2 Method of L-moments

Utilizing the quantile function of ζ_n defined in Eq. (2.43) and (2.46), the relations for the first three L-moments of the RS distribution are derived as,

$$\begin{aligned}
&\lambda_1(\zeta_n) = \gamma + 2\beta R + \alpha(2R)^{1/2}\,\Gamma(3/2),\\
&\lambda_2(\zeta_n) = \beta R + \alpha(R)^{1/2}\left(2^{1/2}-1\right)\Gamma(3/2),\\
&\lambda_3(\zeta_n) = \frac{\beta R}{3} + \alpha(R)^{1/2}\left(2^{1/2}-3+(8/3)^{1/2}\right)\Gamma(3/2).
\end{aligned} \tag{2.52}$$

It can be seen that L-moments are linear functions of the parameters α and β. From these, the empirical estimates of the unknown parameters are estimated as,

$$\begin{aligned}
&\hat{\alpha} = 4.1394\left(l_2(\zeta_n) - 3l_3(\zeta_n)\right)R^{-1/2}\\
&\hat{\beta} = R^{-1}\left(l_2(\zeta_n) - \left(2^{1/2}-1\right)\hat{\alpha}\,\Gamma(3/2)\right)\\
&\hat{\gamma} = l_1(\zeta_n) - 2\hat{\beta}R - \hat{\alpha}(2R)^{1/2}\,\Gamma(3/2)
\end{aligned} \tag{2.53}$$

The relations in Eq. (2.52) can be used for the special RS model with $\alpha = 0$, Eq. (2.47), and the empirical parameters are estimated as,

$$\begin{aligned}
&\alpha = 0,\\
&\hat{\beta} = l_2(\zeta_n)/R,\\
&\hat{\gamma} = l_1(\zeta_n) - 2\hat{\beta}R
\end{aligned} \tag{2.54}$$

II.5 Weibull-Stokes Model

II.5.1 Model development

The Weibull-Stokes (WS) model, as compared with the RS model, is a more general distribution for crests, troughs, and heights of a non-linear system. In this model, it is assumed that the linear process has a Weibull distribution, specifically,

$$f_{\zeta}(x)=\frac{\kappa}{\vartheta}\left(\frac{x}{\vartheta}\right)^{\kappa-1}\exp\left(-\left(\frac{x}{\vartheta}\right)^{\kappa}\right) \qquad x>0 \tag{2.55}$$

Note that Rayleigh distribution, Eq. (2.42), is a special form of the Weibull distribution, Eq. (2.55), with $\kappa=2$ and $\vartheta=(2R)^{1/2}$. Therefore, the WS model has an additional unknown parameter, i.e. the shape parameter κ, which improves the WS model flexibility in capturing the probability distribution of data at the expense of higher model complexity.

Applying the probability distribution of linear process ζ, Eq. (2.55), into the second-order transformation rule defined in Eq. (2.3), it can be shown that the distributions of WS model for $\beta>0$ are,

$$\begin{aligned}
&f_{\zeta_n}(x)=\frac{\kappa}{\vartheta}\left(\frac{\chi-\alpha}{2\beta\vartheta}\right)^{\kappa-1}\chi^{-1}\exp\left(-\left(\frac{\chi-\alpha}{2\beta\vartheta}\right)^{\kappa}\right), \qquad x>\gamma\\
&F_{\zeta_n}(x)=1-\exp\left(-\left(\frac{\chi-\alpha}{2\beta\vartheta}\right)^{\kappa}\right),\\
&x_{\zeta_n}(u)=\gamma+\beta\,\vartheta^2\left(-\ln(1-u)\right)^{2/\kappa}+\alpha\,\vartheta\left(-\ln(1-u)\right)^{1/\kappa}.
\end{aligned} \tag{2.56}$$

For $\beta<0$ the distributions change into,

$$f_{\zeta_n}(x)=\frac{\kappa}{\vartheta\chi}\left[\left(\frac{\chi-\alpha}{2\beta\vartheta}\right)^{\kappa-1}\exp\left(-\left(\frac{\chi-\alpha}{2\beta\vartheta}\right)^{\kappa}\right)H_{\gamma}(x)+\left(\frac{-\chi-\alpha}{2\beta\vartheta}\right)^{\kappa-1}\exp\left(-\left(\frac{-\chi-\alpha}{2\beta\vartheta}\right)^{\kappa}\right)\right],$$

$$F_{\zeta_n}(x)=\left[1-\exp\left(-\left(\frac{\chi-\alpha}{2\beta\vartheta}\right)^{\kappa}\right)\right]H_{\gamma}(x)+\exp\left(-\left(\frac{-\chi-\alpha}{2\beta\vartheta}\right)^{\kappa}\right) \tag{2.57}$$

which are defined for $-\infty<x<\gamma-\frac{\alpha^2}{4\beta}$. The exact analytical form of the quantile function is not available in case of $\beta<0$. Following the same ideology used to simplify the distributions of RS model for $\beta<0$, see section II.4.1, an approximate solution for the WS quantile function for $\beta<0$ is obtained as,

$$x_{\zeta_n}(u)\approx\gamma+\beta\vartheta^2\left(-\ln(1-u)\right)^{2/\kappa}+\alpha\vartheta\left(-\ln(1-u)\right)^{1/\kappa} \tag{2.58}$$

For the WS model with no linear term, i.e. $\alpha=0$, the probability distributions are derived as,

$$f_{\zeta_n}(x)=\frac{\kappa}{2\beta^{1/2}\vartheta(x-\gamma)^{1/2}}\left(\frac{(x-\gamma)^{1/2}}{\beta^{1/2}\vartheta}\right)^{\kappa-1}\exp\left(-\left(\frac{(x-\gamma)^{1/2}}{\beta^{1/2}\vartheta}\right)^{\kappa}\right),\qquad x>\gamma$$

$$F_{\zeta_n}(x)=1-\exp\left(-\left(\frac{(x-\gamma)^{1/2}}{\beta^{1/2}\vartheta}\right)^{\kappa}\right), \tag{2.59}$$

$$x_{\zeta_n}(u)=\gamma+\beta\vartheta^2\left(-\ln(1-u)\right)^{2/\kappa}.$$

II.5.2 Parameter estimation

As shown in Eq.s (2.56) and (2.57), the WS model is a five-parameter distribution. However, the parameter ϑ acts as a normalization parameter for the linear process and its effect can be modeled by the parameters α and β. In that, without loss of generality, it can be assumed that ϑ is known, e.g. $\vartheta = 1$, which makes the WS model a four-parameter distribution. In the following sections, the values of the unknown parameters α, β, γ, and κ are estimated utilizing the method of moments and method of L-moments.

II.5.2.1 Method of moments

Utilizing the quantile function defined in Eq. (2.56), the first four moments of the WS distribution are given in Eq. (2.60). Knowing that $|\beta| \ll \alpha$, the second, third, and fourth moments of ζ_n can be simplified by neglecting the terms with higher powers of β (higher than one) with respect to the other terms. In order to obtain the empirical estimates of the four unknown parameters, the distribution moments given in Eq. (2.60) need to be equated with their corresponding sample statistics. The system of equation then will be solved for the parameters utilizing a non-linear iterative solver. As for the initial values, it can be assumed that $\hat{\kappa}_{initial} = 2.0$, $\hat{\beta}_{initial} = \pm\hat{\alpha}/\alpha'$ where α' is an arbitrary large value, e.g. $\alpha' = 10$. From these, the initial values for the parameters α and γ may be obtained from Eq. (2.50) by substituting R with $\vartheta^2/2$. The relationships given in Eq. (2.60) can be used for the special WS model with $\alpha = 0$ and, furthermore, for empirical parameter estimation.

$$\mu_1(\zeta_n) = \gamma + \frac{2\beta\vartheta^2}{\kappa}\Gamma(2/\kappa) + \frac{\alpha\vartheta}{\kappa}\Gamma(1/\kappa), \tag{2.60}$$

$$\mu_2(\zeta_n) = \frac{4\beta^2\vartheta^4}{\kappa}\left(\Gamma(4/\kappa) - \frac{1}{\kappa}\Gamma(2/\kappa)\right) + \frac{2\alpha\beta\vartheta^3}{\kappa}\left(3\Gamma(3/\kappa) - \frac{2}{\kappa}\Gamma(2/k)\Gamma(1/\kappa)\right) +$$
$$\frac{\alpha^2\vartheta^2}{\kappa}\left(2\Gamma(2/k) - \frac{1}{\kappa}\Gamma^2(1/\kappa)\right),$$

$$\mu_3(\zeta_n) = \frac{4\beta^3\vartheta^6}{\kappa}\left(6\Gamma(6/\kappa) - \frac{24}{\kappa}\Gamma(4/\kappa)\Gamma(2/\kappa) + \frac{16}{\kappa^2}\Gamma^3(2/\kappa)\right) +$$
$$\frac{3\alpha\beta^2\vartheta^5}{\kappa}\left(5\Gamma(5/\kappa) - \frac{12}{\kappa}\Gamma(3/\kappa)\Gamma(2/\kappa) + \frac{8}{\kappa^2}\Gamma(1/\kappa)\Gamma^2(2/\kappa) - \frac{4}{\kappa}\Gamma(4/\kappa)\Gamma(1/\kappa)\right) +$$
$$\frac{3\alpha^2\beta\vartheta^4}{\kappa}\left(4\Gamma(4/\kappa) - \frac{6}{\kappa}\Gamma(3/\kappa)\Gamma(1/\kappa) + \frac{4}{\kappa^2}\Gamma(2/\kappa)\Gamma^2(1/\kappa) - \frac{4}{\kappa}\Gamma^2(2/\kappa)\right) +$$
$$\frac{\alpha^3\vartheta^3}{\kappa}\left(3\Gamma(3/\kappa) - \frac{6}{\kappa}\Gamma(2/\kappa)\Gamma(1/\kappa) + \frac{2}{\kappa^2}\Gamma^3(1/\kappa)\right),$$

$$\mu_4(\zeta_n) = \frac{8\beta^4\vartheta^8}{\kappa}\left(\Gamma(8/\kappa) + \frac{12}{\kappa^2}\Gamma(4/\kappa)\Gamma^2(2/\kappa) - \frac{6}{\kappa}\Gamma(6/\kappa)\Gamma(2/\kappa) - \frac{6}{\kappa^3}\Gamma^4(2/\kappa)\right) +$$
$$\frac{4\alpha\beta^3\vartheta^7}{\kappa}\left(7\Gamma(8/\kappa) + \frac{36}{\kappa^2}\Gamma(3/\kappa)\Gamma^2(2/\kappa) - \frac{30}{\kappa}\Gamma(5/\kappa)\Gamma(2/\kappa) - \frac{6}{\kappa}\Gamma(6/\kappa)\Gamma(1/\kappa) +\right.$$
$$\left.\frac{24}{\kappa^2}\Gamma(4/\kappa)\Gamma(2/\kappa)\Gamma(1/\kappa) - \frac{24}{\kappa^3}\Gamma(2/\kappa)\Gamma(1/\kappa)\right) +$$
$$\frac{6\alpha^2\beta^2\vartheta^6}{\kappa}\left(6\Gamma(6/\kappa) - \frac{16}{\kappa}\Gamma(4/\kappa)\Gamma(2/\kappa) - \frac{10}{\kappa}\Gamma(5/\kappa)\Gamma(1/\kappa) - \frac{12}{\kappa^3}\Gamma^2(2/\kappa)\Gamma^2(1/\kappa) +\right.$$
$$\left.\frac{8}{\kappa^2}\Gamma^3(2/\kappa) + \frac{4}{\kappa^2}\Gamma(4/\kappa)\Gamma^2(1/\kappa) + \frac{24}{\kappa^2}\Gamma(3/\kappa)\Gamma(2/\kappa)\Gamma(1/\kappa)\right) +$$
$$\frac{4\alpha^3\beta\vartheta^5}{\kappa}\left(5\Gamma(5/\kappa) + \frac{9}{\kappa^2}\Gamma(3/\kappa)\Gamma^2(1/\kappa) - \frac{6}{\kappa}\Gamma(3/\kappa)\Gamma(2/\kappa) - \frac{12}{\kappa}\Gamma(4/\kappa)\Gamma(1/\kappa) +\right.$$
$$\left.\frac{6}{\kappa^2}\Gamma(1/\kappa)\Gamma^2(2/\kappa) - \frac{6}{\kappa^3}\Gamma(2/\kappa)\Gamma^3(1/\kappa)\right) +$$
$$\frac{\alpha^4\vartheta^4}{\kappa}\left(4\Gamma(4/\kappa) - \frac{12}{\kappa}\Gamma(3/\kappa)\Gamma(1/\kappa) + \frac{12}{\kappa^2}\Gamma(2/\kappa)\Gamma^2(1/\kappa) - \frac{3}{\kappa^3}\Gamma^4(2/\kappa)\right).$$

II.5.2.2 Method of L-moments

The first four L-moments of the non-linear process ζ_n in the WS model are derived as,

$$\lambda_1(\zeta_n) = \gamma + \frac{2\beta\vartheta^2}{\kappa}\Gamma(2/\kappa) + \frac{\alpha\vartheta}{\kappa}\Gamma(1/\kappa), \tag{2.61}$$

$$\lambda_2(\zeta_n) = \frac{2\beta\vartheta^2}{\kappa}\left(1-(1/2)^{2/\kappa}\right)\Gamma(2/\kappa) + \frac{\alpha\vartheta}{\kappa}\left(1-(1/2)^{1/\kappa}\right)\Gamma(1/\kappa),$$

$$\lambda_3(\zeta_n) = \frac{2\beta\vartheta^2}{\kappa}\left(1-3(1/2)^{2/\kappa}+2(1/3)^{2/\kappa}\right)\Gamma(2/\kappa) + \frac{\alpha\vartheta}{\kappa}\left(1-3(1/2)^{1/\kappa}+2(1/3)^{1/\kappa}\right)\Gamma(1/\kappa),$$

$$\lambda_4(\zeta_n) = \frac{2\beta\vartheta^2}{\kappa}\left(1-6(1/2)^{2/\kappa}+10(1/3)^{2/\kappa}-5(1/4)^{2/\kappa}\right)\Gamma(2/\kappa) + \frac{\alpha\vartheta}{\kappa}\left(1-6(1/2)^{1/\kappa}+10(1/3)^{1/\kappa}-5(1/4)^{1/\kappa}\right)\Gamma(1/\kappa)$$

Similar to relations in Eq. (2.52) of the RS model L-moments, L-moments of the WS model are linear functions of the parameters α and β, however, the L-moments in the WS model are non-linear functions of the Weibull parameters κ and ϑ. In order to obtain the empirical estimates of the four unknown parameters with the method of L-moments, one needs to equate the L-moments given in Eq. (2.61) with their corresponding unbiased sample estimator. To obtain the solutions, a non-linear iterative solver is required. As initial values for the iterations, it could be assumed that

$\kappa_{initial} = 2.0$ and the initial estimates of the other three parameters may be obtained from the RS parameter estimates given in Eq. (2.53) and substitution of R with $\vartheta^2/2$.

II.6 Extreme Analysis

Designers tasked to study the wave-structure interaction problems are routinely involved in evaluating their design models in extreme environmental conditions. For this purpose, they evaluate the behavior of the offshore structures in extreme design sea-states with 100yr – 1000yr return periods. Furthermore, within the design sea-state, they are mostly interested in extreme phenomena, e.g. crest maxima. Here, the probability distribution of the crest maxima $\zeta_{\max}$ is studied utilizing the probability distribution of non-linear crests ζ_n obtained from RS and WS models.

Assuming that crests ζ_n are independent identically distributed (i.i.d) random variables with PDF f_{ζ_n} and CDF F_{ζ_n}, the PDF $f_{\zeta_{\max}}$ and CDF $F_{\zeta_{\max}}$ of the crest maxima $\zeta_{\max}$ in N waves can be obtained from the ordered value statistics theory [45] as,

$$f_{\zeta_{\max}}(x) = N f_{\zeta_n}(x)\left[F_{\zeta_n}(x)\right]^{N-1} \tag{2.62}$$

$$F_{\zeta_{\max}}(x) = \left[F_{\zeta_n}(x)\right]^{N} \tag{2.63}$$

For large number of waves N,

$$\lim_{N\to\infty} F_{\zeta_{\max}}(x) = \lim_{N\to\infty}\left[F_{\zeta_n}(x)\right]^{N} = \begin{cases} 1, & F_{\zeta_n} = 1 \\ 0, & F_{\zeta_n} < 1 \end{cases} \tag{2.64}$$

which indicates that the limit distributions are degenerate. To avoid degeneracy, one needs to look for a linear transformation such that,

$$\lim_{N\to\infty}\left[F_{\zeta_n}(x)\right]^N = F_{\zeta_{\max}}\left(\frac{x-a_N}{b_N}\right) \tag{2.65}$$

and a_N and b_N are extreme constants and depend on N. The only non-degenerate family of distributions that satisfies Eq. (2.65) is the generalized extreme value distribution (GEVD). The GEVD reduces to well-known inverse Weibull, Gumbel, and Fréchet families of maxima probability distributions for GEVD shape parameter κ_{GV} of $\kappa_{GV}>0$, $\kappa_{GV}=0$, and $\kappa_{GV}<0$, respectively. A necessary and sufficient condition for F_{ζ_n} to belong to the maximal domain of attraction is that,

$$\lim_{\varepsilon\to 0}\frac{x_{\zeta_n}(1-\varepsilon)-x_{\zeta_n}(1-2\varepsilon)}{x_{\zeta_n}(1-2\varepsilon)-x_{\zeta_n}(1-4\varepsilon)} = 2^{-\kappa_{GV}} \tag{2.66}$$

It can be shown that Eq. (2.66) limits to 1.0 for both the RS and WS quantile functions, defined in Eq.s (2.43) and (2.56), respectively. This indicates that in both models the F_{ζ_n} belongs to the Gumbel maximal domain of attraction and the asymptotic probability distribution of $\zeta_{\max}$ may be represented as,

$$F_{\zeta_{\max}}(x) = \exp\left(-\exp\left(-\frac{x-a_N}{b_N}\right)\right) \qquad -\infty < x < +\infty \tag{2.67}$$

and consequently estimates of a_N and b_N may be obtained from,

$$a_N = x_{\zeta_n}\left(1-\frac{1}{N}\right), \qquad b_N = x_{\zeta_n}\left(1-\frac{1}{Ne}\right) - a_N. \tag{2.68}$$

where, $e \approx 2.7183$ is the Euler's number. Applying these equations, the extreme parameters for RS model are estimated as,

$$\begin{aligned} a_N &= \gamma + 2\beta R \ln(N) + \alpha \left(2R \ln(N)\right)^{1/2}, \\ b_N &= 2\beta R + \alpha (R)^{1/2} \left(2 \ln(N)\right)^{-1/2}. \end{aligned} \tag{2.69}$$

Similarly for WS model the parameters a_N and b_N can be obtained from,

$$\begin{aligned} a_N &= \gamma + \beta \vartheta^2 \left(\ln(N)\right)^{2/\kappa} + \alpha \vartheta \left(\ln(N)\right)^{1/\kappa}, \\ b_N &= \beta \vartheta^2 \left[\left(\ln(N)+1\right)^{2/\kappa} - \left(\ln(N)\right)^{2/\kappa}\right] + \alpha \vartheta \left[\left(\ln(N)+1\right)^{1/\kappa} - \left(\ln(N)\right)^{1/\kappa}\right]. \end{aligned} \tag{2.70}$$

Assuming that $\zeta_{\max}$ follows the Gumbel probability distribution, the first three moments of $\zeta_{\max}$ are obtained in the form of,

$$\begin{aligned} \mu_1(\zeta_{\max}) &= a_N + b_N \gamma_{EM}, \\ \mu_2(\zeta_{\max}) &= \frac{\pi^2}{6} b_N^2, \\ \mu_3(\zeta_{\max}) &= 2\zeta_R(3) b_N^3. \end{aligned} \tag{2.71}$$

where $\gamma_{EM} \approx 0.5772$ is the Euler-Mascheroni constant and $\zeta_R(z)$ is the Riemann zeta function that is $\zeta_R(3) \approx 1.2021$ at $z = 3$. The first three L-moments of $\zeta_{\max}$ are derived as,

$$\begin{aligned} \lambda_1(\zeta_{\max}) &= a_N + b_N \gamma_{EM}, \\ \lambda_2(\zeta_{\max}) &= b_N \ln(2), \\ \lambda_3(\zeta_{\max}) &= b_N \left(2\ln(3) - 3\ln(2)\right). \end{aligned} \tag{2.72}$$

CHAPTER III

MODEL UNCERTAINTY ANALYSIS

III.1 Introduction

Uncertainty analysis is a crucial part of every model development study. The analysis basically evaluates the reliability of the model estimates and also the sensitivity of the model results to the initial conditions and assumptions. As described in the previous chapter, the structural form of the semi-empirical probability distribution is developed based on a mathematical model. The difference between the theoretical probability distribution and the semi-empirical probability distribution is in the model parameter estimation procedure. In the theoretical approach, the parameter values are obtained based on some theoretical assumptions and relating the model parameters to the basic characteristics of the random process. In the semi-empirical approach, however, the parameters are directly related to the sample statistics, e.g. sample L-moments, and sample moments. In the derivation of the probability distribution, it is assumed that the parameters are deterministic constants while in reality even the theoretically estimated parameters contain a level of uncertainty. Besides, in case of empirical parameter estimation, the sample variability causes uncertainty of the parameter estimates. Additionally, the sample quality is always a concern in the empirical parameter estimation.

In the first part of this chapter, the effects of the model parameter uncertainty on the semi-empirical model statistics are studied. The model parameter uncertainty is a

common concern of both empirically and theoretically estimated parameters. In the second part of this chapter, the effects of sample variability on the statistics of the semi-empirical models are studied utilizing Monte-Carlo simulations. Additionally, the sample size effect on the model estimates is evaluated and the performance of the parameter estimation methods, i.e. method of moments (MoM) and method of L-moments (MoL), is compared.

Faulty measurements and glitches are a major concern in the empirical parameter estimation. However, because of the unknown nature and source of these errors, it is almost impossible to model their effects precisely. Our experience on different sample data sets in the field of ocean engineering/oceanography indicates that the errors mostly happen on the extreme observations and cause contaminated tail distributions. Here, the sensitivity of MoM and MoL to the faulty measurements in the tail of the distribution is studied using illustrative examples.

In this study, bias, variance, and root-mean-squared error (RMSE) are utilized to quantify the uncertainty of an estimate $\hat{\varphi}$ of the true value φ. Bias Δ_{φ} is defined as the difference between the parameter's expected value $E[\hat{\varphi}] = \mu_{\hat{\varphi}}$ and the true value φ, specifically,

$$\Delta_{\varphi} = E[\hat{\varphi}] - \varphi \tag{3.1}$$

The positive bias indicates an overestimation while negative bias shows the opposite. Variance is an indication of variability around the mean value which for a sample of size N is estimated from,

$$\sigma_{\hat{\varphi}}^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} \left(\hat{\varphi}_i - \mu_{\hat{\varphi}} \right)^2 \tag{3.2}$$

Another useful statistic is the standard deviation $\sigma_{\hat{\varphi}}$ and that is the square-root of variance. RMSE is an indication of the difference between the model estimates and the true value, and can be estimated from,

$$RMSE(\hat{\varphi}) = \left(E\left[\left(\varphi - \hat{\varphi} \right)^2 \right] \right)^{1/2} = \left(\Delta_{\varphi}^2 + \sigma_{\hat{\varphi}}^2 \right)^{1/2} \tag{3.3}$$

The RMSE may be normalized by the true value to obtain the percentage RMSE.

III.2 Model Parameter Uncertainty

In this section the effects of the model parameters variability on the quantile function of the distribution models are evaluated. For illustrative purposes, numerical examples are utilized. The study is performed for the Gaussian-Stokes (GS), Rayleigh-Stokes (RS), and Weibull-Stokes (WS) distribution models.

III.2.1 Gaussian-Stokes model

The quantile function of the non-linear process ζ_n in the GS model is given in Eq. (2.31). In the initial model, it was assumed that the three model parameters, i.e. α, β and γ are deterministic constants, while in general, these parameters can be modeled as random variables with expected value μ_1 and variance σ^2, e.g. $\left(\mu_1(\alpha), \sigma_{\alpha}^2 \right)$. From Eq. (2.31), the expected value and variance of quantile with probability of u are obtained in terms of the model parameter statistics as,

$$\mu_1\left(x_{\zeta_n}(u)\right) = \mu_1(\gamma) + \mu_1(\beta)\left(\Phi^{-1}(u)\right)^2 + \mu_1(\alpha)\Phi^{-1}(u)$$
$$\sigma^2_{x_{\zeta_n}(u)} = \sigma^2_\gamma + 2\left(\Phi^{-1}(u)\right)^2\sigma^2_{\gamma,\beta} + 2\Phi^{-1}(u)\sigma^2_{\gamma,\alpha} + \sigma^2_\beta\left(\Phi^{-1}(u)\right)^4 + 2\left(\Phi^{-1}(u)\right)^3\sigma^2_{\alpha,\beta} + \sigma^2_\alpha\left(\Phi^{-1}(u)\right)^2 \quad (3.4)$$

where $\sigma^2_{y,z}$ is the covariance of the random variables y and z. Assuming that, α, β and γ are mutually independent random variables, the equation for the variance $\sigma^2_{x_{\zeta_n}(u)}$ simplifies to,

$$\sigma^2_{x_{\zeta_n}(u)} = \sigma^2_\gamma + \sigma^2_\beta\left(\Phi^{-1}(u)\right)^4 + \sigma^2_\alpha\left(\Phi^{-1}(u)\right)^2 \quad (3.5)$$

As shown in Eq.(3.5), the contribution of σ^2_γ to $\sigma^2_{x_{\zeta_n}}$ is constant and does not vary with u, while, the contributions of σ^2_β and σ^2_α are respectively amplified by the factors $\left(\Phi^{-1}(u)\right)^4$ and $\left(\Phi^{-1}(u)\right)^2$ that are both varying with the probability u. Variation of these amplification factors with the exceedance probability $P = 1-u$ is presented in Figure 2. In this figure, the amplification factors are only plotted for $u \geq 0.5$ and the distributions are symmetric around $u = 0.5$. As shown in Figure 2, the parameters variability is considerably magnified on the distribution tails while their effects are minimized at $u = 0.5$. Comparing the distributions given in Figure 2, one can easily see that the variability of β has higher influence on the extreme statistics.

For illustrative purposes, the variability of the GS quantile function for an arbitrarily selected set of parameter statistics is presented in Figure 3. In evaluation of the distribution in this figure, Monte-Carlo simulation technique is utilized from which 100,000 independent samples are generated. The samples are generated assuming that

the parameters α, β, and γ are mutually independent, α has a lognormal distribution with $\mu_1(\alpha) = 1.0$, and $\sigma_\alpha = 0.05$, β has a lognormal distribution with $\mu_1(\beta) = 0.1$, and $\sigma_\beta = 0.05$, γ has a Gaussian distribution with $\mu_1(\gamma) = 0.0$, and $\sigma_\gamma = 0.05$. The solid line in Figure 3 shows the mean quantile value averaged over the Monte-Carlo samples and the dashed lines show the 95 percent confidence levels. As shown in Eq. (3.4) the mean (expected value) distribution is independent of the parameters' variance. As explained earlier and shown in Figure 3, the variability of the GS quantile increases on the tails of the distribution and is minimum at the distribution median $u = 0.5$. The log-normally distributed α and β have caused skewed quantile distribution at the probability of u. This can be realized from the asymmetrical confidence levels around the mean value shown in Figure 3. Although the variability of the parameters is fairly small in this example, the variability of the extreme values is considerably high. This indicates the sensitivity of the extreme statistics to the model parameters and justifies the application of more robust parameter estimation methods.

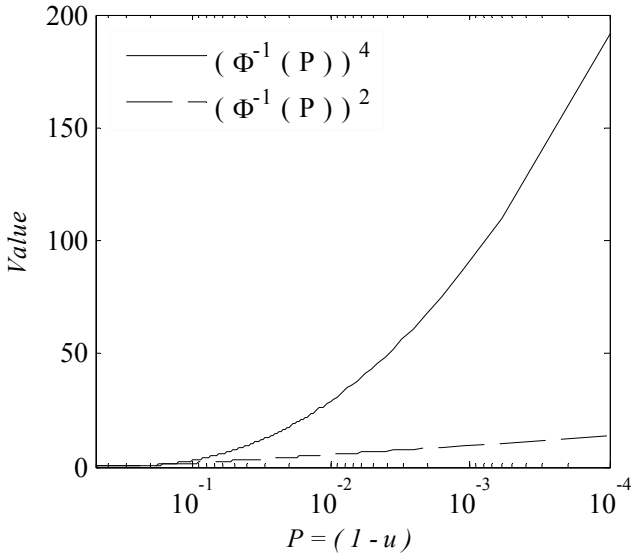


Figure 2 Contribution of the parameters variance in the quantile uncertainty of the Gaussian-Stokes model.

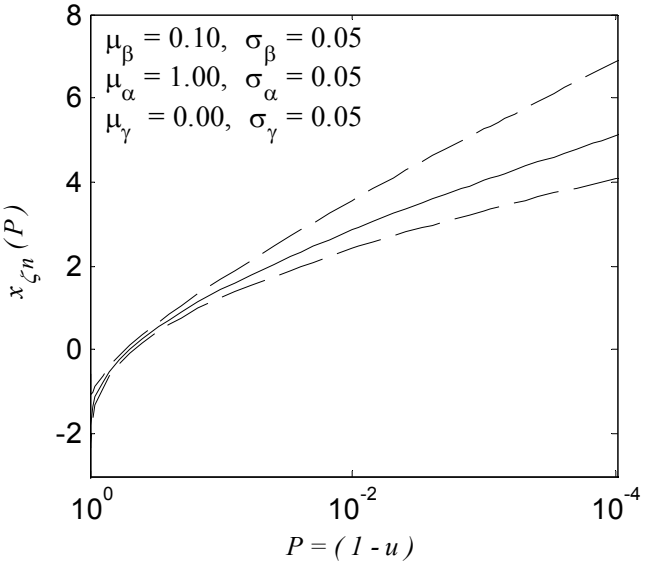


Figure 3 Effect of model parameters uncertainty on the quantiles of the Gaussian-Stokes model. Solid line represents the mean value and the dashed lines show the 95% confidence levels.

III.2.2 Rayleigh-Stokes Model

Similarly to what was done for the GS model, the expected value and variance of the quantile function $x_{\zeta_n}(u)$ of the RS model, Eq. (2.43), are derived in terms of model parameters statistics,

$$\begin{aligned}&\mu_1\left(x_{\zeta_n}(u)\right)=\mu_1(\gamma)+\mu_1(\beta)U_R{}^2+\mu_1(\alpha)U_R,\\&\sigma^2_{x_{\zeta_n}(u)}=\sigma^2_\gamma+2U_R{}^2\sigma^2_{\gamma,\beta}+2U_R\sigma^2_{\gamma,\alpha}+\sigma^2_\beta U_R{}^4+2U_R{}^3\sigma^2_{\alpha,\beta}+\sigma^2_\alpha U_R{}^2.\end{aligned} \tag{3.6}$$

where $U_R=\left(-2R\ln(1-u)\right)^{1/2}$ is the quantile function of the linear process ζ. Assuming that α, β, and γ are mutually independent random variables, the variance of $x_{\zeta_n}(u)$ simplifies into the summation of the contribution of the parameter variances, specifically,

$$\sigma^2_{x_{\zeta_n}(u)}=\sigma^2_\gamma+\sigma^2_\beta U_R{}^4+\sigma^2_\alpha U_R{}^2 \tag{3.7}$$

As shown here, the contribution of variability of γ is constant throughout the probability u, while the contributions of σ^2_β and σ^2_α are amplified with factors varying with u, i.e. U^2 and U, respectively. The distribution of these amplification factors with respect to the exceedance probability $P=1-u$ is presented in Figure 4. As can be seen in this figure, the contribution of parameter uncertainties increases on the distribution tail, and the uncertainty in β has higher effects on the variability of the tail as compared to the uncertainty in α.

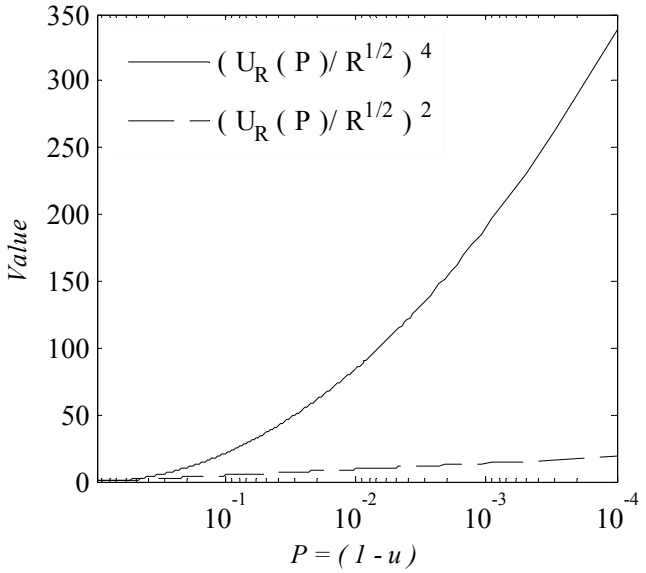


Figure 4 Contribution of the parameters variance in the quantile uncertainty of the Rayleigh-Stokes semi-empirical model.

The variability of the RS quantile function is estimated for an arbitrarily selected set of parameter statistics and the results are presented in Figure 5. For this purpose 100,000 samples are generated utilizing Monte-Carlo simulations, assuming that parameters are mutually independent, α has a lognormal distribution with $\mu_1(\alpha)=1.0$, and $\sigma_\alpha=0.05$, β has a lognormal distribution with $\mu_1(\beta)=0.1$, and $\sigma_\beta=0.05$, γ has a Gaussian distribution with $\mu_1(\gamma)=0.0$, and $\sigma_\gamma=0.05$, and $R=1.0$ is constant. Similar to Figure 3, the solid line in Figure 5 indicates the mean value and the dashed lines indicate the 95-percent confidence level of the quantile function. Regarding the first relation in Eq. (3.6), the quantile mean distribution is independent of the covariance of the model parameters. As discussed earlier and shown in Figure 5, the quantile uncertainty increases on the distribution tail and even with relatively small variability of

the parameters the variability of the extreme values is significant. The selection of a skewed distribution for α and β, i.e. lognormal, has resulted in a skewed distribution for $x_{\zeta_n}(u)$ which can be recognized from the asymmetrical confidence levels around the mean value, shown in Figure 5.

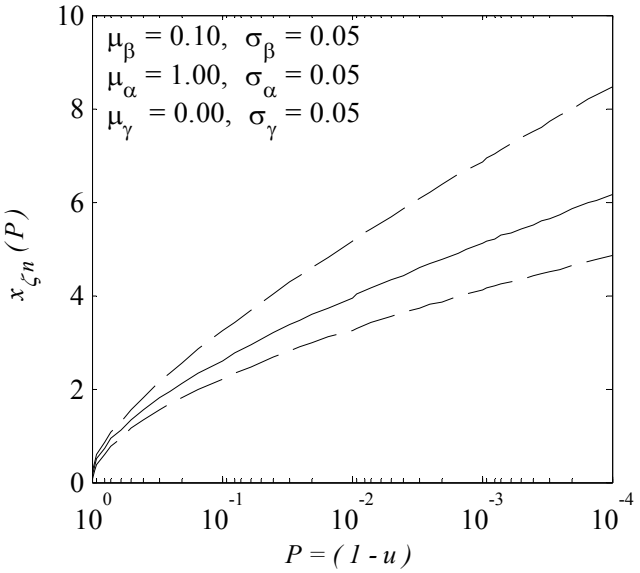


Figure 5 Effect of model parameters uncertainty on the quantiles of the Rayleigh-Stokes model. Solid line represents the mean value and the dashed lines show the 95% confidence levels.

As discussed in the previous chapter, the statistics of the response maxima $\zeta_{\max}$ can be obtained from the probability distribution of the non-linear process ζ_n. For large number of waves N, the expected maxima $E[\zeta_{\max}] = \mu_1(\zeta_{\max})$ can be derived explicitly in terms of the model parameters α, β, and γ, see Eq. (2.71). From that, one may obtain the mean and standard deviation of the $E[\zeta_{\max}]$ in N waves as,

$$\mu_1\left(E\left[\zeta_{\max}\right]\right) = \mu_1(\gamma) + \mu_1(\beta) N_{1R} + \mu_1(\alpha) N_{2R}$$
$$\sigma^2_{E[\zeta_{\max}]} = \sigma^2_{\gamma} + N^2_{1R}\sigma^2_{\beta} + N^2_{2R}\sigma^2_{\alpha} + N_{1R}\sigma^2_{\gamma,\beta} + N_{2R}\sigma^2_{\gamma,\alpha} + N_{1R}N_{2R}\sigma^2_{\alpha,\beta} \tag{3.8}$$

where $N_1 = 2R\left(\ln(N) + \gamma_{EM}\right)$ and $N_2 = (R)^{1/2}\left(\left(2\ln(N)\right)^{1/2} + \gamma_{EM}\left(2\ln(N)\right)^{-1/2}\right)$.

For independent model parameters, the covariance terms on the right hand side of $\sigma^2_{E(\zeta_{\max})}$ become zero. Utilizing the same parameter statistics used in development of Figure 5, the PDF of $E\left[\zeta_{\max}\right]$ for $N = 1{,}000$ is presented in Figure 6. $E\left[\zeta_{\max}\right]$ in this example has a mean and variance of $\mu_1\left(E\left[\zeta_{\max}\right]\right) = 5.37$ and $\sigma^2_{E[\zeta_{\max}]} = 0.61$ with 95-percent lower and upper confidence levels of 4.30 and 7.31, respectively. The relatively high variance of the expected maxima and the wide range of the confidence levels clearly show the sensitivity of the extreme statistics to the random variability of the model parameters.

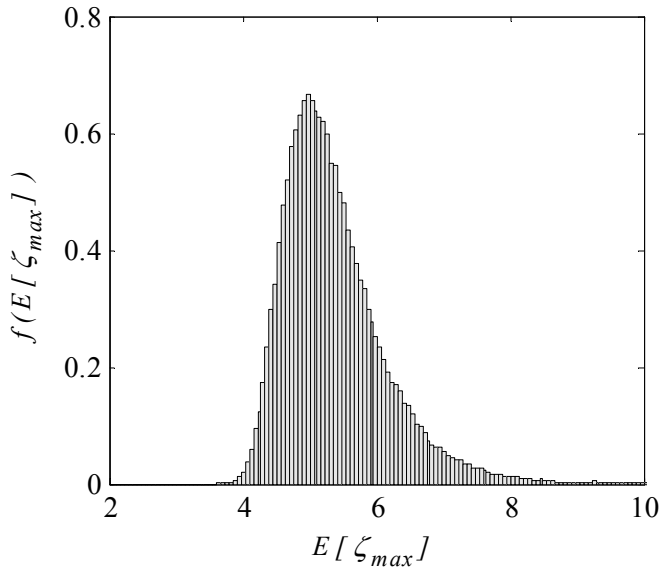


Figure 6 Probability density function of the expected maxima in the Rayleigh-Stokes model.

III.2.3 Weibull-Stokes model

The uncertainty analysis is more complicated in the case of the WS semi-empirical model since the quantile function is a non-linear function of the shape parameter κ which itself can be a random variable. For this purpose, the quantile function of the WS model, Eq.(2.56), is re-written as,

$$x_{\zeta_n}(u) = \gamma + \beta U_\kappa^2 + \alpha U_\kappa \tag{3.9}$$

where, $U_\kappa = \vartheta\left(-\ln(1-u)\right)^{1/\kappa}$. Note that here all the four parameters α, β, γ and U_κ are random variables. From this, the mean and variance of $x_{\zeta_n}(u)$ are obtained as,

$$\begin{aligned} &\mu_1\left(x_{\zeta_n}(u)\right) = \mu_1(\gamma) + \mu_1\left[\beta U_\kappa{}^2\right] + \mu_1\left[\alpha U_\kappa\right], \\ &\sigma^2_{x_{\zeta_n}(u)} = \sigma^2_\gamma + 2\sigma^2_{\gamma,\beta U_\kappa{}^2} + 2\sigma^2_{\gamma,\alpha U_\kappa} + \sigma^2_{\beta U_\kappa{}^2} + 2\sigma^2_{\alpha U_\kappa,\beta U_\kappa{}^2} + \sigma^2_{\alpha U_\kappa}. \end{aligned} \tag{3.10}$$

For mutually independent random variables Eq. (3.10) simplifies to,

$$\begin{aligned} &\mu_1\left(x_{\zeta_n}(u)\right) = \mu_1(\gamma) + \mu_1(\beta)\,\mu_1\left(U_\kappa{}^2\right) + \mu_1(\alpha)\,\mu_1(U_\kappa), \\ &\sigma^2_{x_{\zeta_n}(u)} = \sigma^2_\gamma + \sigma^2_{\beta U_\kappa{}^2} + \sigma^2_{\alpha U_\kappa}. \end{aligned} \tag{3.11}$$

As can be seen in Eq.s (3.10) and (3.11), the mean value $\mu_1\left(x_{\zeta_n}(u)\right)$ varies with the variance of the shape parameter because of the term $\mu_1\left(U_\kappa{}^2\right)$. Utilizing 100,000 samples generated from Monte-Carlo simulation, the variability of $x_{\zeta_n}(u)$ is presented in Figure 7 for an arbitrarily selected set of parameter statistics, i.e. α has a lognormal distribution with $\mu_1(\alpha) = 1.0$ and $\sigma_\alpha = 0.05$, β has a lognormal distribution with $\mu_1(\beta) = 0.1$ and $\sigma_\beta = 0.05$, γ has a normal distribution with $\sigma_\gamma = 0.05$ and

$\mu_1(\gamma) = 0.0$, and κ has a lognormal distribution with $\mu_1(\kappa) = 2.0$ and $\sigma_\kappa = 0.05$, and $\vartheta = \sqrt{2}$ is constant. The solid line in Figure 7 indicates the mean value and the dashed lines represent the 95-percent confidence levels. The parameters of this example are chosen in a way that the mean distribution shown in Figure 7 is identical to what was estimated for RS model in Figure 5. Comparing these two figures, one might think that the additional random variability of the shape parameter κ in the WS model has not affected the distributions significantly. However, this is mainly because of the small standard deviation of κ in this example. In Figure 8, the uncertainty of the WS quantile with respect to the standard deviation of the shape parameter σ_κ is presented while the statistics of other parameters are kept the same. The center lines in Figure 8 are the mean distribution while the lines on the sides represent the distributions of the 95-percent confidence levels. As mentioned earlier and shown in this figure, quantile mean value varies with σ_κ. The distributions given in Figure 8 indicate that the uncertainty of the shape parameter has higher contribution to the variability of the extreme values.

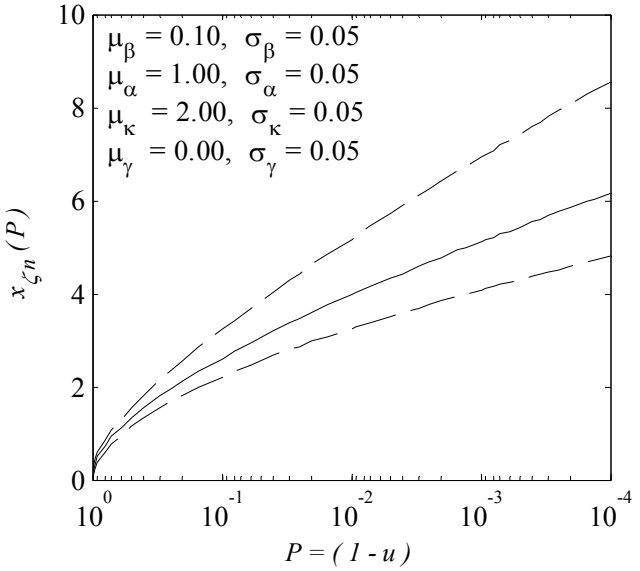


Figure 7 Effect of the model parameters uncertainty on the quantiles of the Weibull-Stokes model. Solid line represents the mean value and the dashed lines show the 95% confidence levels.

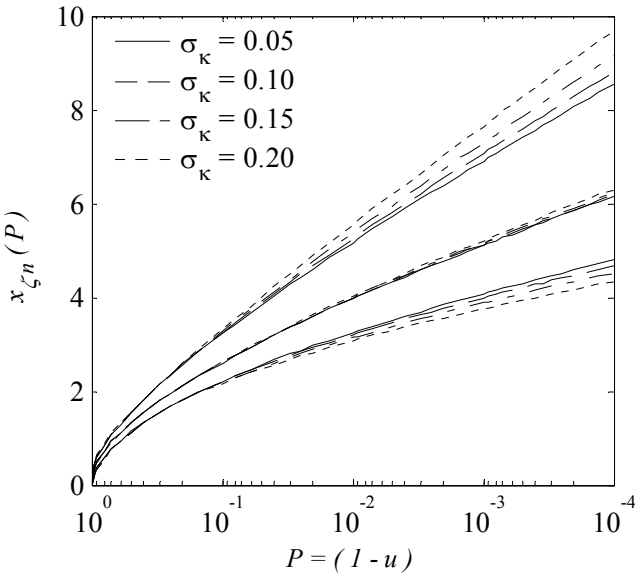


Figure 8 Variation of Weibull-Stokes quantile statistics with the shape parameter standard deviation. Central lines represent the mean value and the outer lines show the 95% confidence levels.

The effects of model parameter uncertainty on the statistics of the extreme statistics are studied in more detail. Particularly, the uncertainty of the expected maxima caused by the WS model parameter uncertainty is evaluated. Utilizing the relation given in (2.71) for the expected maxima in the WS model, the mean and standard deviation of expected maxima can be obtained from,

$$\begin{aligned} \mu_1\left(E\left[\zeta_{\max}\right]\right) &= \mu_1\left(a_N\right) + \gamma_{EM}\,\mu_1\left(b_N\right) \\ \sigma^2_{E[\zeta_{\max}]} &= \sigma^2_{a_N} + \gamma^2_{EM}\sigma^2_{b_N} + \gamma_{EM}\sigma^2_{a_N,b_N} \end{aligned} \tag{3.12}$$

where the relation for mean and variance of the asymptotic extreme parameters a_N and b_N can be found using their relation with the model parameters given in Eq. (2.70). Similar to Figure 6, the PDF of $E\left[\zeta_{\max}\right]$ for $N = 1,000$ is given in Figure 9 for the same parameter statistics used in Figure 7. $E\left[\zeta_{\max}\right]$ in this example has a mean and variance of $\mu_1\left(E\left[\zeta_{\max}\right]\right) = 5.37$ and $\sigma^2_{E[\zeta_{\max}]} = 0.64$ with 95-percent lower and upper confidence limits of 4.26 and 7.34, respectively. The confidence limits are slightly wider than the ones estimated for the RS model with similar parameter statistics and constant shape parameter. Note that the higher uncertainty of the shape parameter not only increases the variance of the expected maxima but also increases the mean value which is equivalent to an increase of the bias in the model estimates. In Figure 10, the distributions of $\mu_1\left(E\left[\zeta_{\max}\right]\right)$ and the 95-percent confidence levels as a function of σ_κ are presented. In estimation of these distribution, it is assumed that the variability of the other three parameters α, β, and γ is zero. The distributions in Figure 10 clearly show the sensitivity of the expected maxima to the uncertainty of the shape parameter.

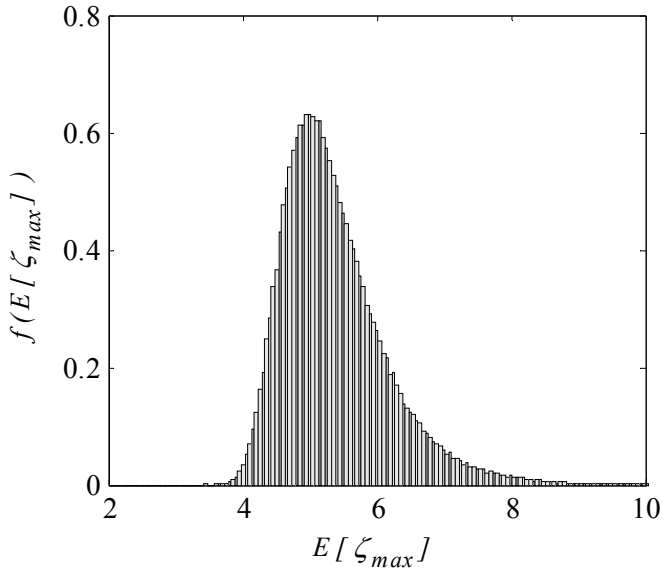


Figure 9 Probability density function of the expected maxima of the Weibull-Stokes model.

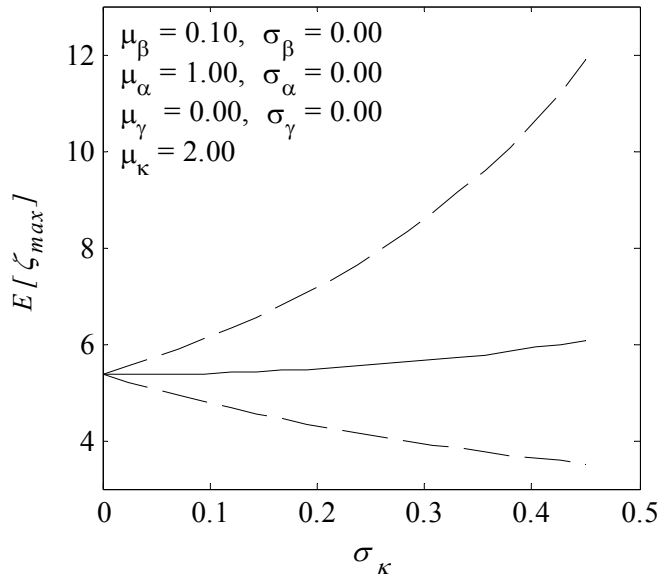


Figure 10 Variation of the expected maxima statistics with the variability of the shape parameter in the Weibull-Stokes model. Solid line represents the mean value and dashed lines show the 95% confidence levels.

III.3 Sample Variability and Parameter Estimation Uncertainty

In order to study the effects of sample variability, 25,000 samples of size N_s are generated from the model quantile function with known parameters. The sample moments and L-moments and consequently the model parameters are estimated for each sample with method of moments (MoM) and method of L-moments (MoL). The variability of the sample moments and L-moments, model parameters, quantiles, and in case of RS and WS models, extreme statistics are evaluated. Our analysis indicates that there is no clear correlation between the initial model parameter values and parameter estimation uncertainty. In that, the results of analysis for only a set of parameters statistics are presented. To evaluate the sample size effect, samples with $N_s = 2{,}000$, 1,000, 330, and 100 observations are studied which respectively approximate the total number of waves in 3hr, 1hr, 30min, and 17min measured data.

III.3.1 Moments vs. L-moments

In Table 3, the uncertainty of the moments and L-moments of a GS model with $\alpha = 1.0, \beta = 0.1,$ and $\gamma = 0.0$ is presented. In this table the true value represents the distribution statistics, bias Δ, standard deviation σ, and RMSE are estimated utilizing Eq.s (3.1), Eq. (3.2), and (3.3), respectively, and the RMSEs are normalized with the proper true value to obtain the percentage RMSE.

As shown in Table 3, the bias in the moments and L-moments of the GS model is considerably small even for a small sample size of $N_s = 100$. As expected, the variability of the sample estimates increases as the sample size decreases and

consequently the percentage RMSE is larger in smaller samples. As can be seen in Table 3, higher order moments are more sensitive to the sample size as compared to the corresponding L-moments. The standard deviation of sample second and third L-moments remain fairly small as compared to the standard deviation of the second and third moments. It has been seen that, despite the first moment and L-moment which are identical by definition, sample L-moments are more efficient than ordinary moments even for large samples.

Table 3 Uncertainty analysis of the moments and L-moments of the Gaussian-Stokes model with model parameters $\alpha = 1.0$, $\beta = 0.1$, and $\gamma = 0.0$.

Estimate	N_s	True Value	σ	Δ	RMSE %	Estimate	True Value	σ	Δ	RMSE %
$l_1(\zeta_n)$	2000	0.10	0.02	0.00	22.67	$\hat{\mu}_1(\zeta_n)$	0.10	0.02	0.00	22.67
	1000		0.03	0.00	31.82			0.03	0.00	31.82
	330		0.06	0.00	55.51			0.06	0.00	55.51
	100		0.10	0.00	100.80			0.10	0.00	100.80
$l_2(\zeta_n)$	2000	0.56	0.01	0.00	1.68	$\hat{\mu}_2(\zeta_n)$	1.02	0.04	0.00	3.50
	1000		0.01	0.00	2.38			0.05	0.00	4.98
	330		0.02	0.00	4.14			0.09	0.00	8.62
	100		0.04	0.00	7.56			0.16	0.00	15.80
$l_3(\zeta_n)$	2000	0.06	0.01	0.00	10.40	$\hat{\mu}_3(\zeta_n)$	0.61	0.08	0.00	13.95
	1000		0.01	0.00	14.81			0.12	0.00	20.03
	330		0.01	0.00	25.71			0.21	0.00	34.60
	100		0.03	0.00	48.04			0.39	0.00	64.78

Similar to Table 3, in Table 4 the uncertainty of the moments and L-moments of the RS model with $\alpha = 1.0, \beta = 0.1, \gamma = 0.0$, and $R = 1.0$ are presented. Similar conclusions can be made from this table. As shown in Table 4, moments and L-moments

of the RS model remain unbiased even for small samples. Once more, sample moments found to be more sensitive to the sample size and the standard deviation of the second and third moments are considerably higher as compared with the variability of the corresponding L-moments. The results obtained for $N_s = 2000$ and $N_s = 1000$ are not significantly different while the error in the statistics of smaller samples is considerably increased.

Table 4 Uncertainty analysis of the moments and L-moments of the Rayleigh-Stokes model with model parameters $\alpha = 1.0, \beta = 0.1, \gamma = 0.0$, and $R = 1.0$.

Estimate	Ns	True Value	σ	Δ	RMSE %	Estimate	True Value	σ	Δ	RMSE %
$l_1(\zeta_n)$	2000	1.45	0.02	0.00	1.31	$\hat{\mu}_1(\zeta_n)$	1.45	0.02	0.00	1.31
	1000		0.03	0.00	1.85			0.03	0.00	1.85
	330		0.05	0.00	3.19			0.05	0.00	3.19
	100		0.07	0.00	5.15			0.07	0.00	5.15
$l_2(\zeta_n)$	2000	0.47	0.01	0.00	1.81	$\hat{\mu}_2(\zeta_n)$	0.72	0.03	0.00	3.95
	1000		0.01	0.00	2.54			0.04	0.00	5.55
	330		0.02	0.00	4.44			0.07	0.00	9.66
	100		0.03	0.00	6.84			0.09	0.00	12.77
$l_3(\zeta_n)$	2000	0.08	0.01	0.00	6.88	$\hat{\mu}_3(\zeta_n)$	0.56	0.07	0.00	12.29
	1000		0.01	0.00	9.70			0.10	0.00	17.19
	330		0.01	0.00	16.85			0.17	0.00	29.83
	100		0.02	0.00	26.33			0.18	0.00	31.75

Similar to Table 3 and Table 4, the results of uncertainty analysis on the moments and L-moments of a WS model with parameters $\alpha = 1.0, \beta = 0.1, \gamma = 0.0$, $\kappa = 2.0$, and $\vartheta = \sqrt{2}$ are presented in Table 5. The model parameters are selected in a way that the statistics of the WS model are equal to the statistics of the RS model studied

earlier. For this example, the uncertainty of the first four moments and L-moments are estimated. The additional unknown parameter, shape parameter κ, adds to the complexity of this distribution model while increasing the model flexibility in capturing the probability distribution of data. Similar conclusions as of RS and GS model can be made for the first three moments and L-moments of WS models. Comparing the results of uncertainty analysis of the RS and WS models given respectively in Table 4 and Table 5, it can be seen that the errors in the sample estimates are consistently higher in the WS model. This is caused by the more complex structure of the WS model. A more interesting observation is obtained by comparing the uncertainty of the fourth L-moment and moment in Table 5. As shown in this table, fourth L-moment of the WS model is an unbiased sample estimator while the fourth moment is positively biased and the bias is higher in smaller samples. Additionally, the fourth moment is found to be highly sensitive to the sample size. The RMSE of the third and fourth L-moments are not significantly different while the RMSE of the fourth moment is constantly higher than the RMSE of the third moment.

Table 5 Uncertainty analysis of the moments and L-moments of the Weibull-Stokes model with model parameters $\alpha = 1.0, \beta = 0.1$, $\gamma = 0.0$, $\kappa = 2.0$, and $\vartheta = \sqrt{2}$.

Estimate	Ns	True Value	σ	Δ	RMSE %	Estimate	True Value	σ	Δ	RMSE %
$l_1(\zeta_n)$	2000	1.45	0.02	0.00	1.32	$\hat{\mu}_1(\zeta_n)$	1.45	0.02	0.00	1.32
	1000		0.03	0.00	1.83			0.03	0.00	1.83
	330		0.05	0.00	3.24			0.05	0.00	3.24
	100		0.09	0.00	5.87			0.09	0.00	5.87
$l_2(\zeta_n)$	2000	0.47	0.01	0.00	1.79	$\hat{\mu}_2(\zeta_n)$	0.72	0.03	0.00	3.89
	1000		0.01	0.00	2.54			0.04	0.00	5.54
	330		0.02	0.00	4.44			0.07	0.00	9.68
	100		0.04	0.00	8.05			0.13	0.00	17.56
$l_3(\zeta_n)$	2000	0.08	0.01	0.00	6.77	$\hat{\mu}_3(\zeta_n)$	0.56	0.07	0.00	12.06
	1000		0.01	0.00	9.73			0.10	0.00	17.12
	330		0.01	0.00	16.92			0.17	0.00	29.91
	100		0.02	0.00	30.98			0.31	0.00	55.22
$l_4(\zeta_n)$	2000	0.06	0.00	0.00	6.42	$\hat{\mu}_4(\zeta_n)$	2.10	0.29	0.01	14.02
	1000		0.01	0.00	9.21			0.42	0.01	19.76
	330		0.01	0.00	16.20			0.73	0.03	34.72
	100		0.02	0.00	30.06			1.43	0.09	67.97

III.3.2 Parameter estimates

In the previous section, the uncertainty of the moments and L-moments of the semi-empirical models were studied. The sample moments and L-moments are respectively considered as the sample estimators for the MoM and MoL empirical parameter estimation methods. In Table 6, the uncertainty of the empirically estimated parameters of the GS model is evaluated. In this table, the efficiency of the two-parameter estimation methods is compared. As presented in Table 6, the bias in the estimated parameters is almost zero for both parameter estimation methods even in small

samples. This was expected as the sample moments and L-moments of the GS model, as shown in Table 3, are unbiased sample estimators. The performance of the parameter estimation methods is comparable while the RMSE of the parameters estimated by method of L-moments is slightly less. It is observed that γ has the highest uncertainty as compared to the other model parameters, i.e. α and β.

Table 6 Uncertainty analysis of the empirical parameter estimation for the Gaussian-Stokes model.

Estimate	N_s	True Value	Method of L-moments				Method of Moments			
			σ	Δ	RMSE	RMSE %	σ	Δ	RMSE	RMSE %
$\hat{\alpha}$	2000	1.00	0.02	0.00	0.02	1.68	0.02	0.00	0.02	1.68
	1000		0.02	0.00	0.02	2.38	0.02	0.00	0.02	2.38
	330		0.04	0.00	0.04	4.14	0.04	0.00	0.04	4.12
	100		0.08	0.00	0.08	7.56	0.08	-0.01	0.08	7.53
$\hat{\beta}$	2000	0.10	0.01	0.00	0.01	10.40	0.01	0.00	0.01	11.97
	1000		0.01	0.00	0.01	14.81	0.02	0.00	0.02	17.14
	330		0.03	0.00	0.03	25.71	0.03	0.00	0.03	29.31
	100		0.05	0.00	0.05	48.04	0.05	0.00	0.05	52.82
$\hat{\gamma}$	2000	0.00	0.02	0.00	0.02	NA	0.03	0.00	0.03	NA
	1000		0.03	0.00	0.03	NA	0.04	0.00	0.04	NA
	330		0.06	0.00	0.06	NA	0.06	0.00	0.06	NA
	100		0.11	0.00	0.11	NA	0.11	0.00	0.11	NA

Similarly, the results of parameter uncertainty analysis of the RS model are presented in Table 7. It is important to notice that in this model the estimates of the MoM are slightly biased especially the estimates of the small samples. The estimates of the MoL, however, remain unbiased even for small samples. The variability of the parameter estimates of both methods is comparable while the parameters estimated with

MoL have slightly smaller standard deviation and RMSE. In the RS model, $\hat{\alpha}$ has the highest variability, while because of the small true value of $\beta = 0.1$, β has the highest percentage RMSE. As observed here, the parameters of RS model are more sensitive to the sample size and have higher uncertainty as compared to the parameters of the GS model.

Table 7 Uncertainty analysis of the empirical parameter estimation for the Rayleigh-Stokes model.

Estimate	N_s	True Value	Method of L-moments				Method of Moments			
			σ	Δ	RMSE	RMSE %	σ	Δ	RMSE	RMSE %
$\hat{\alpha}$	2000	1.00	0.05	0.00	0.05	5.27	0.07	0.00	0.07	6.92
	1000		0.07	0.00	0.07	7.40	0.10	0.01	0.10	9.64
	330		0.13	0.00	0.13	12.89	0.16	0.01	0.16	16.27
	100		0.21	0.00	0.21	21.17	0.23	0.03	0.23	22.90
$\hat{\beta}$	2000	0.10	0.02	0.00	0.02	21.38	0.03	0.00	0.03	26.39
	1000		0.03	0.00	0.03	30.10	0.04	0.00	0.04	36.71
	330		0.05	0.00	0.05	52.31	0.06	-0.01	0.06	61.97
	100		0.08	0.00	0.08	83.01	0.09	-0.01	0.09	86.09
$\hat{\gamma}$	2000	0.00	0.03	0.00	0.03	NA	0.04	0.00	0.04	NA
	1000		0.04	0.00	0.04	NA	0.05	0.00	0.05	NA
	330		0.06	0.00	0.06	NA	0.09	-0.01	0.09	NA
	100		0.11	0.00	0.11	NA	0.12	-0.01	0.12	NA

The parameter uncertainty is more critical in case of the WS model, as this model is a four-parameter distribution and in addition to the first three moments (L-moments), the estimate of the fourth moment (L-moment) is required for parameter estimation. Table 8 presents the uncertainty of the empirically estimated WS model parameters. In general, the application of the fourth moment in the parameter estimation causes more

variability in the parameter estimates as compared to the parameter estimates of the RS model. As shown in Table 8 for large samples $N_s = 2000, 1000$, the performance of the parameter estimation methods, i.e. MoM and MoL, is similar while the uncertainty of the parameters estimated by MoL is constantly less. However, in case of small samples, $N_s = 330, 100$, the estimates of MoM are highly biased, and the variability of these estimates is excessively high. Comparing the results of the parameter estimation methods for small samples, one can see that MoL performs significantly better in these examples.

Table 8 Uncertainty analysis of the empirical parameter estimation for the Weibull-Stokes model.

Estimate	N_s	True Value	Method of L-moments				Method of Moments			
			σ	Δ	RMSE	RMSE %	σ	Δ	RMSE	RMSE %
$\hat{\alpha}$	2000	1.00	0.11	-0.02	0.11	11.11	0.09	0.01	0.09	9.30
	1000		0.12	-0.01	0.13	12.50	0.14	0.02	0.14	14.18
	330		0.16	0.01	0.16	15.62	1.41	-0.13	1.41	141.14
	100		0.22	0.01	0.22	21.81	4.94	-1.28	5.11	510.79
$\hat{\beta}$	2000	0.10	0.09	0.01	0.09	87.93	0.09	-0.02	0.09	90.49
	1000		0.10	0.01	0.10	96.62	0.12	-0.03	0.12	119.01
	330		0.11	-0.01	0.11	113.11	0.67	0.04	0.67	674.01
	100		0.14	-0.03	0.14	140.45	2.27	0.59	2.34	2342.78
$\hat{\gamma}$	2000	0.00	0.04	0.01	0.05	NA	0.07	0.03	0.07	NA
	1000		0.06	0.02	0.06	NA	0.08	0.04	0.09	NA
	330		0.09	0.04	0.09	NA	0.68	0.10	0.69	NA
	100		0.14	0.06	0.16	NA	2.26	0.12	2.27	NA
$\hat{\kappa}$	2000	2.00	0.24	0.00	0.24	12.24	0.27	-0.11	0.29	14.43
	1000		0.28	-0.03	0.28	13.91	0.34	-0.14	0.36	18.13
	330		0.34	-0.09	0.35	17.69	0.94	-0.06	0.94	46.92
	100		0.47	-0.16	0.50	24.86	2.62	0.61	2.69	134.61

III.3.3 Quantile estimates and extreme statistics

Ocean engineers are mainly concerned about the uncertainty of the models final output, which in case of the semi-empirical models is the probability distribution. In this section the effect of sample variability on the quantile distribution of GS, RS, and WS semi-empirical models is studied. The Monte-Carlo simulations are utilized here to estimate the quantiles with different probabilities. Moreover, in case of RS and WS models, the variability of the expected maxima in 1000 waves ($N = 1,000$) is evaluated.

In Figure 11 and Figure 12 the effect of sample variability on the GS model quantiles estimated respectively by MoL and MoM is studied. In these figures, the performance of the MoM and MoL parameter estimation methods for sample sizes of $N_s = 2,000$, 1,000, 330, 100 are evaluated. The solid lines in these figures indicate the true distributions used for Monte-Carlo simulation, the heavy dashed line is the mean quantile distribution estimated from averaging over 25,000 Monte-Carlo samples, and the light dashed lines indicate the 95-percent confidence limits. The difference between the solid and heavy dashed lines represents the bias in the estimates and the difference between confidence limits indicates the variability of the estimates. As shown in Figure 11 and Figure 12, the bias in the quantile estimates of both MoM and MoL is fairly small. The variability of quantile estimates is larger on the tails of the distributions, as expected, and increases for the smaller samples. Comparing the distributions shown in Figure 11 and Figure 12, it is observed that the results of MoM and MoL are fairly close in the studied cases.

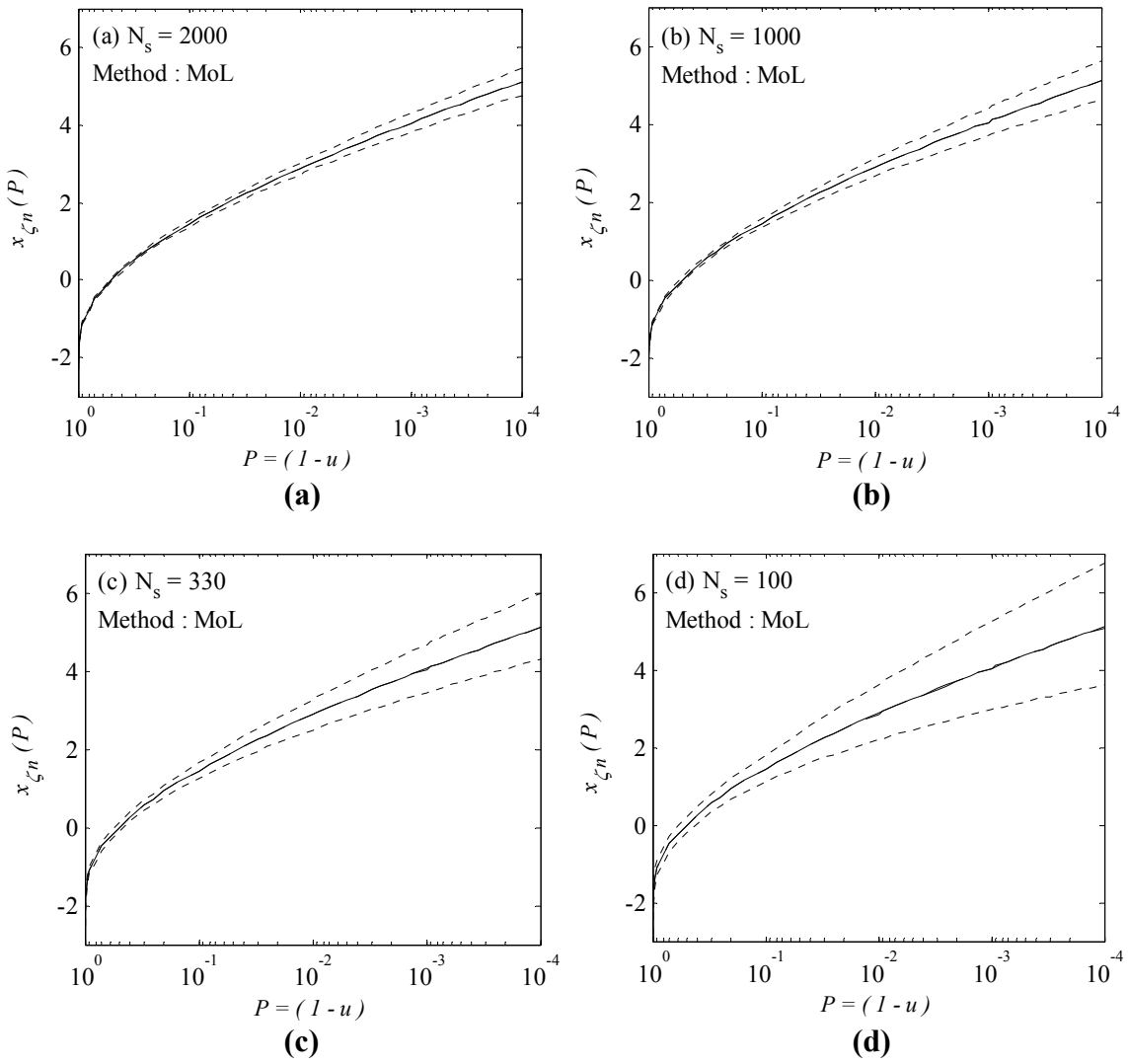


Figure 11 Effect of sample variability on the quantile estimates of the Gaussian-Stokes model estimated from method of L-moments. Solid line represents the true value, heavy dashed line represents the mean value, and dashed lines represent the lower and upper 95% confidence levels.

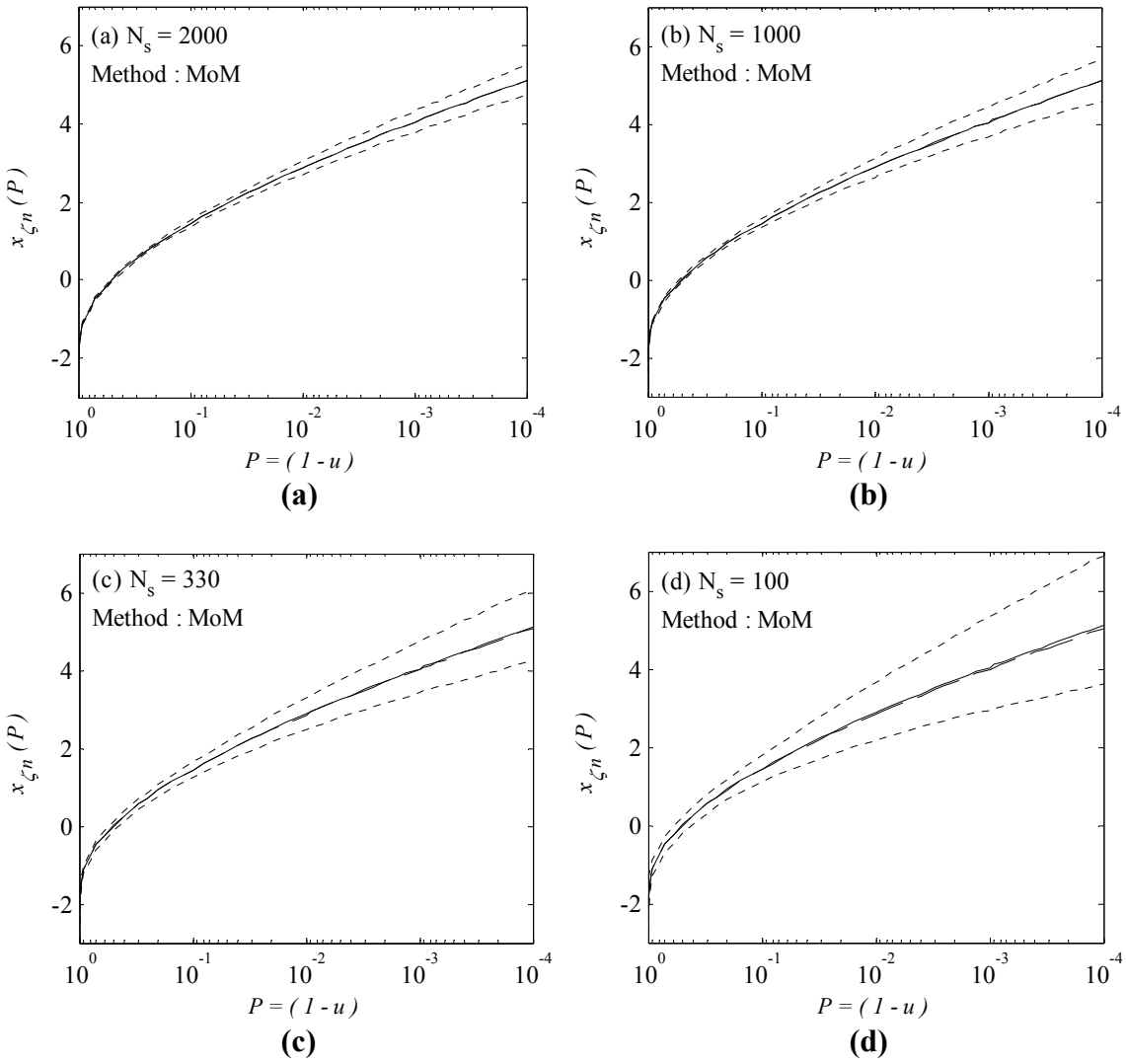


Figure 12 Effect of sample variability on the quantile estimates of the Gaussian-Stokes model estimated from method of moments. Solid line represents the true value, heavy dashed line represents the mean value, and dashed lines represent the lower and upper 95% confidence levels.

Similarly, in Figure 13 and Figure 14 the effect of sample variability on the RS model quantiles estimated respectively by MoL and MoM is studied. As shown in these figures, the quantile estimates of MoL remain unbiased even for small samples. However, small sample estimates of MoM under-predict the true quantile with small probability of exceedances (see Figure 14 (d)). The variability of the estimates of MoL and MoM are comparable for almost all of the studied cases. This can be realized by comparing the confidence levels obtained from the two methods. The confidence levels found to be almost symmetric around the mean level indicating that the probability distribution of the quantile with probability of u is not significantly skewed.

The results of quantile uncertainty analysis of the WS model are presented in Figure 15 and Figure 16 for estimates obtained from MoL and MoM, respectively. As shown in these figures, both MoL and MoM tend to underestimate the extreme quantiles while the absolute bias in the MoM estimates is constantly higher. As expected, the quantile estimates are highly sensitive to the sample size and the estimates of small samples are significantly uncertain. The confidence levels are skewed towards the smaller values (skewed to left). This is more sensible in the quantile estimates of the small samples (see e.g. Figure 15 (d) and Figure 16 (d)). Similar to what we had for GS and RS models, the variability of the quantile estimates of MoL and MoM found to be close. The results of this analysis indicate that although the WS model has more flexibility in capturing the probability distribution of data, its estimates of small samples are highly uncertain and should be used with more care.

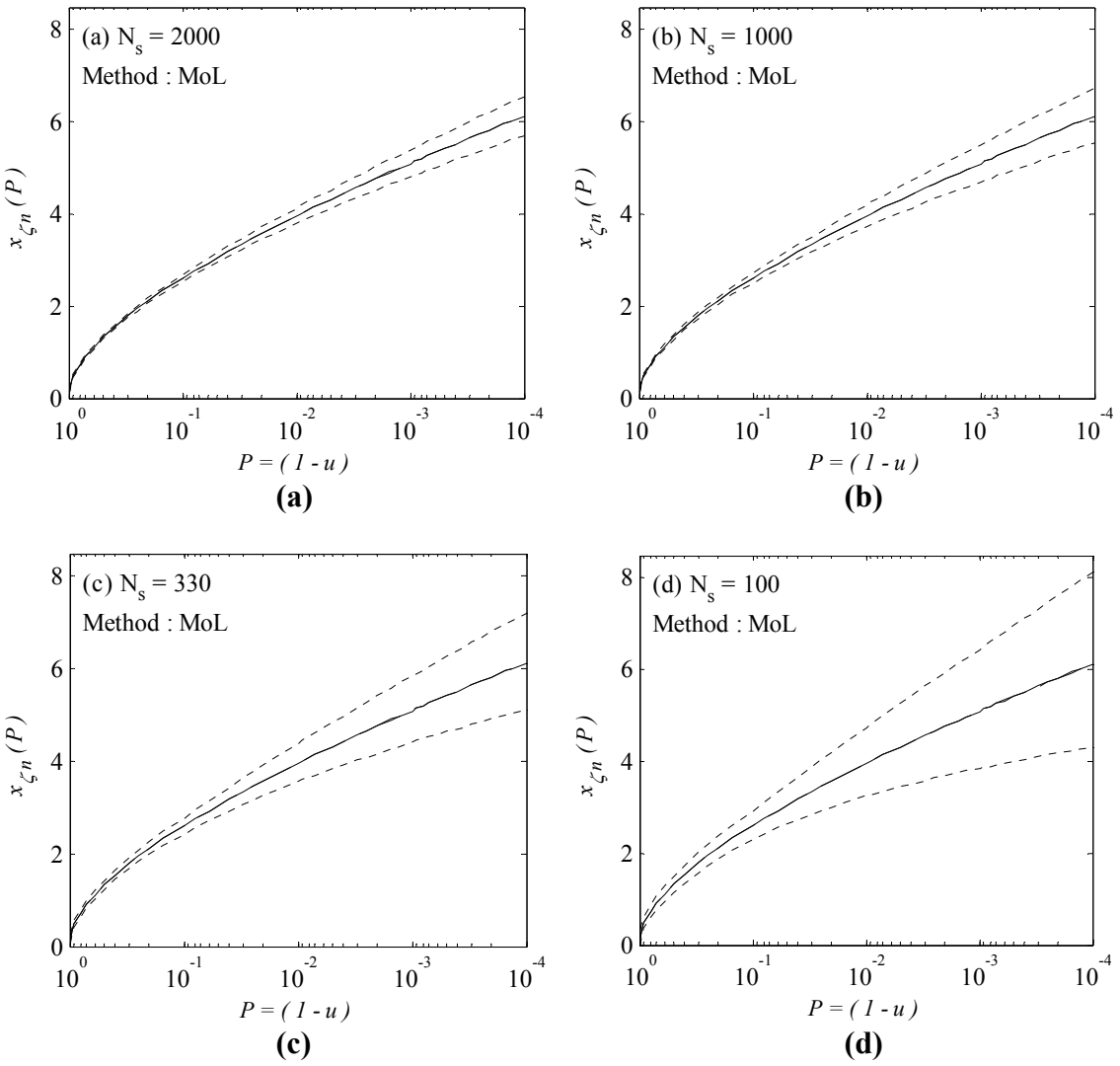


Figure 13 Effect of sample variability on the quantile estimates of the Rayleigh-Stokes model estimated from method of L-moments. Solid line represents the true value, heavy dashed line represents the mean value, and dashed lines represent the lower and upper 95% confidence levels.

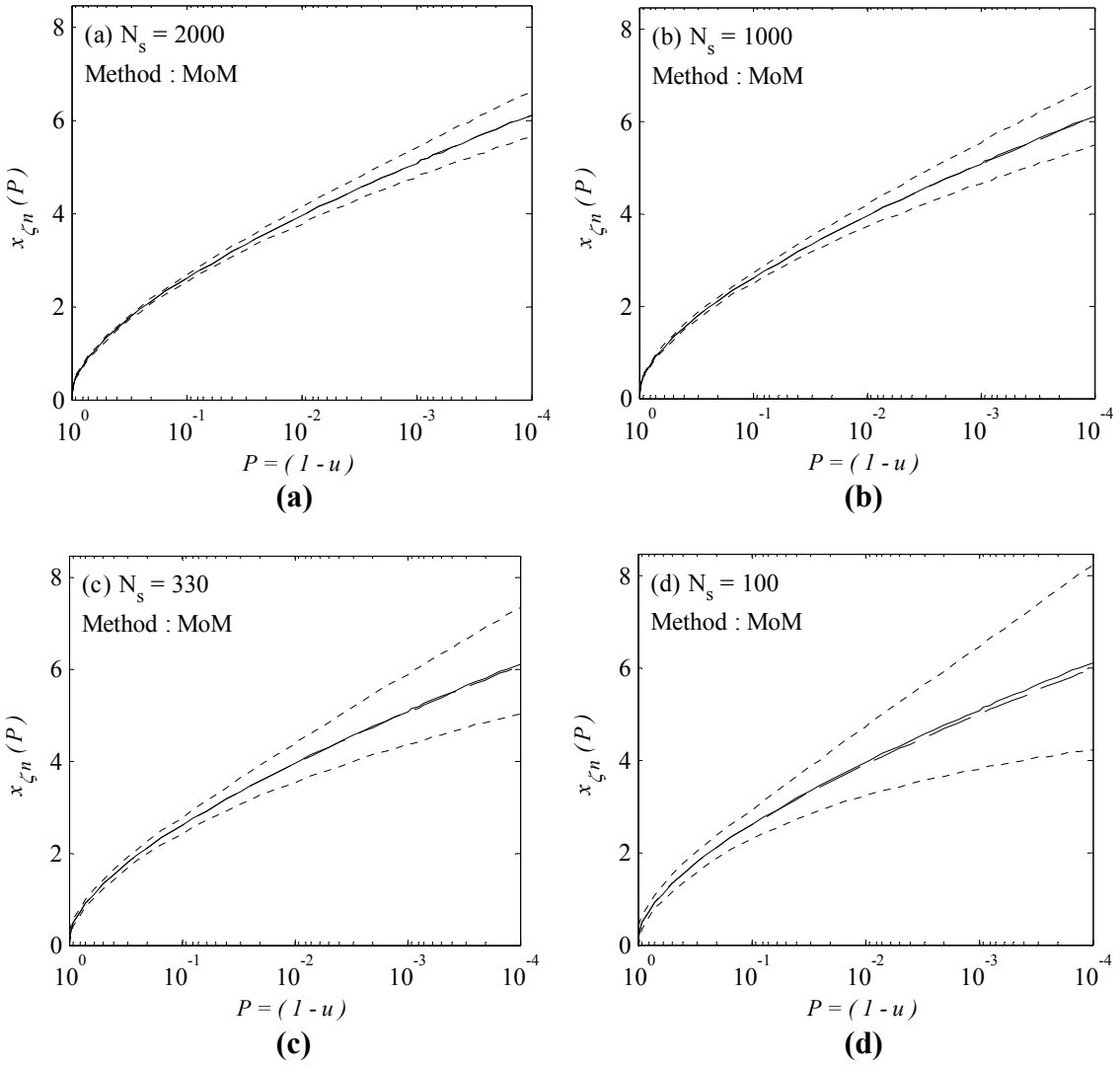


Figure 14 Effect of sample variability on the quantile estimates of the Rayleigh-Stokes model estimated from method of moments. Solid line represents the true value, heavy dashed line represents the mean value, and dashed lines represent the lower and upper 95% confidence levels.

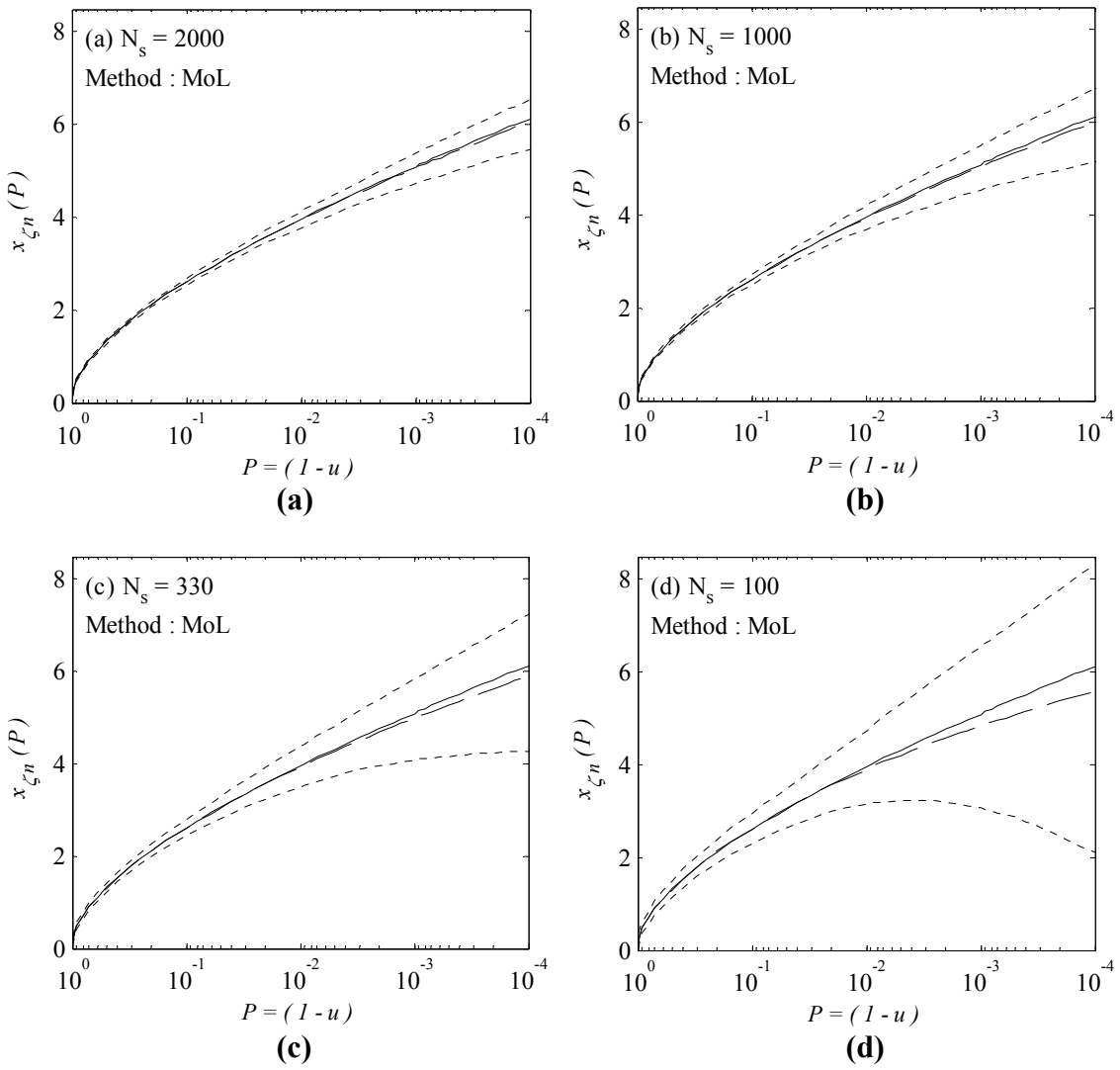


Figure 15 Effect of sample variability on the quantile estimates of the Weibull-Stokes model estimated from method of L-moments. Solid line represents the true value, heavy dashed line represents the mean value, and dashed lines represent the lower and upper 95% confidence levels.

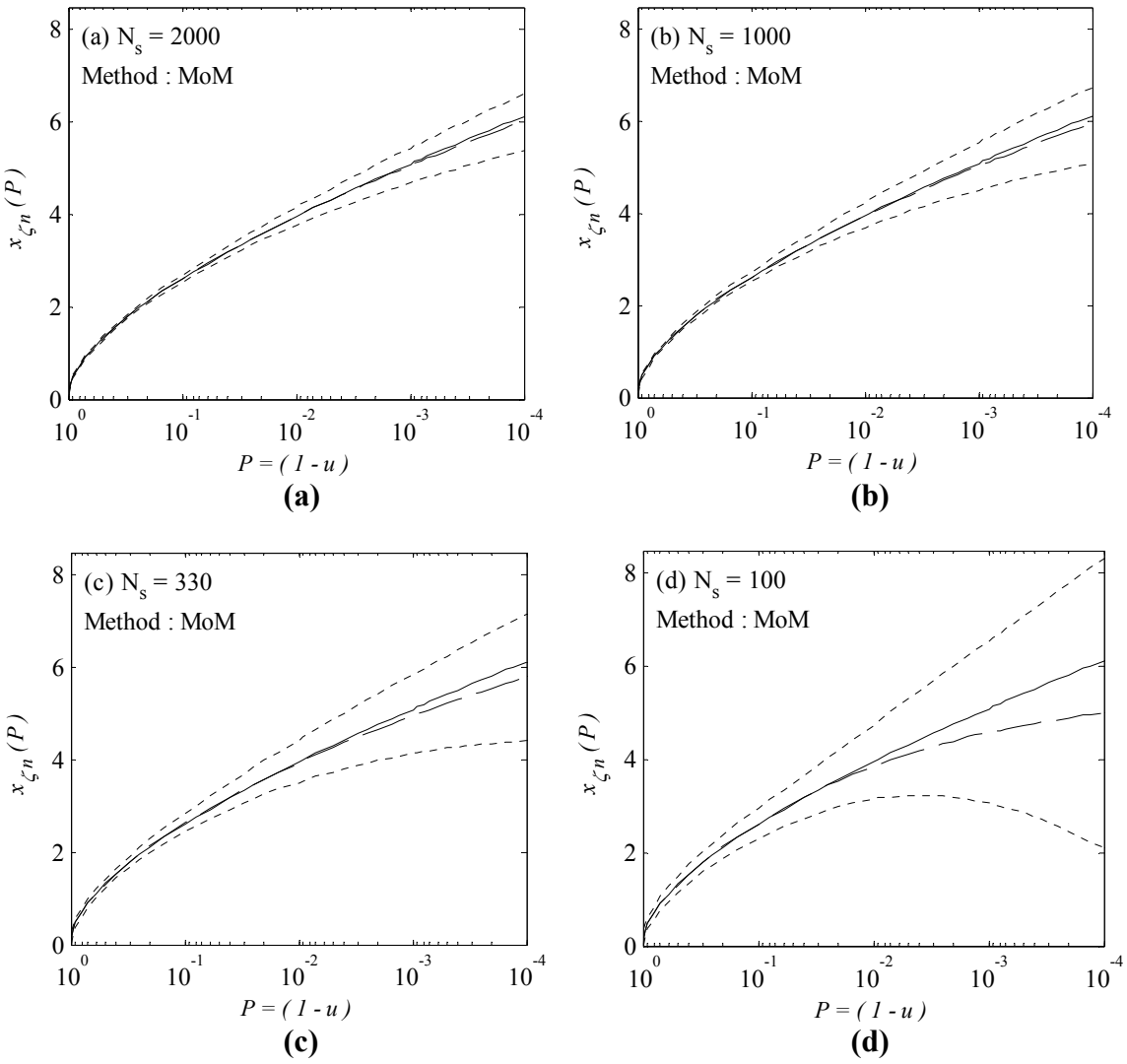


Figure 16 Effect of sample variability on the quantile estimates of the Weibull-Stokes model estimated from method of moments. Solid line represents the true value, heavy dashed line represents the mean value, and dashed lines represent the lower and upper 95% confidence levels.

In Table 9 and Table 10, the results of uncertainty analysis on the expected maxima of RS and WS models are presented, respectively. In these tables, the performance of the parameter estimation methods MoL and MoM are evaluated. As shown in Table 9, the expected maxima estimates of MoL are almost unbiased while MoM under-predicts the true value. The absolute bias in the estimates of MoM is higher for smaller samples. The standard deviation of the MoM and MoL estimates are relatively close. As expected, the uncertainty of the expected maxima of WS is higher as compared to corresponding statistics of the RS model which can be realized by comparing the statistics given in Table 9 and Table 10. As shown in Table 10 for WS model, both MoL and MoM tend to underestimate the expected maxima while the absolute bias is significantly higher in the estimates of the MoM.

The values given in Table 9 and Table 10 can be used for design purposes. Particularly, for a specific sample size one could get an idea about the level of uncertainty of the model estimates. Besides, while designing an experiment, one could decide about the required sample size by specifying the desired level of uncertainty.

Table 9 Uncertainty analysis of the expected maxima of the Rayleigh-Stokes model.

Estimate	Ns	True Value	Method of L-moments			Method of Moments		
			σ	Δ	RMSE %	σ	Δ	RMSE %
$\hat{E}[\zeta_{max}]$	2000	5.37	0.16	0.00	3.07	0.18	-0.01	3.44
	1000		0.23	0.00	4.33	0.26	-0.01	4.79
	330		0.40	0.00	7.52	0.44	-0.03	8.14
	100		0.74	-0.01	13.79	0.76	-0.11	14.31

Table 10 Uncertainty analysis of the expected maxima of the Weibull-Stokes model.

Estimate	Ns	True Value	Method of L-moments			Method of Moments		
			σ	Δ	RMSE %	Σ	Δ	RMSE %
$\hat{E}[\zeta_{\max}]$	2000	5.37	0.19	-0.04	3.67	0.22	-0.05	4.26
	1000		0.28	-0.07	5.40	0.31	-0.09	5.97
	330		0.51	-0.14	9.90	0.53	-0.19	10.55
	100		1.01	-0.33	19.88	1.11	-0.83	21.93

III.4 Effects of Faulty Extreme Measurements

Faulty measurements and glitches are more probable to happen on the tail part of the distribution where the data sets are sparse. Here, two major issues, i.e. unexpectedly large observations and flat tail are studied. The unexpectedly large values may be caused by measurement noise or glitches and their main characteristic is that they do not follow the general pattern of the data points. The flat tail usually happens because of instrumentation limitation where the measurement system cannot measure observations larger (or smaller) than a certain value. The difficulty here is the recognition of the faulty measurements from actual extreme values. There is evidence in the literature that the extremely large (or small) events may not be originated from the same physical process as that of the medium or small size events. For instance, extremely large ocean wave crests are expected to be more non-linear than small and medium size wave crests and other physical processes, e.g. wave breaking and whitecapping, may affect the distribution of these extreme events.

The main purpose of this study is to evaluate the effect of contaminated distribution tail on the estimates of MoM and MoL. For this purpose numerical examples are presented in which the efficiency of the parameter estimation methods is evaluated on samples with artificially contaminated tails. The examples consist of samples with unexpectedly large values and flat tails generated from RS semi-empirical distribution utilizing Monte-Carlo simulation. The samples with unexpectedly large values are generated by adding extra uncertainty to the N_t largest sample points of the initial sample. The additional error is generated from the positive side of a Gaussian distribution (half Gaussian) with $\mu_1(e)$ and σ_e. It is assumed that error magnitude is ascending towards the larger sample points. In case of samples with flat tail, it is assumed that the N_t largest sample points of the initial sample have a constant magnitude equal to x_{N_t-1}. For a better understanding, in Figure 17 the distributions of two sample data sets with $N_s = 1,000$ observations and (a) unexpectedly large values and (b) flat tail are presented. The samples are generated from RS model with $\alpha = 1.0$, $\beta = 0.1$, $\gamma = 0.0$, and $R = 1.0$. In the sample with unexpectedly large values, the additional error is generated from a standard Gaussian distribution with $\mu_1(e) = 0.0$ and $\sigma_e = 0.2$ and is added to the $N_t = 50$ largest observations. For the sample with flat tail, it is assumed that the measurement system is not able to measure observations over $x_{\zeta_n}(u = 0.05) = 3.05$.

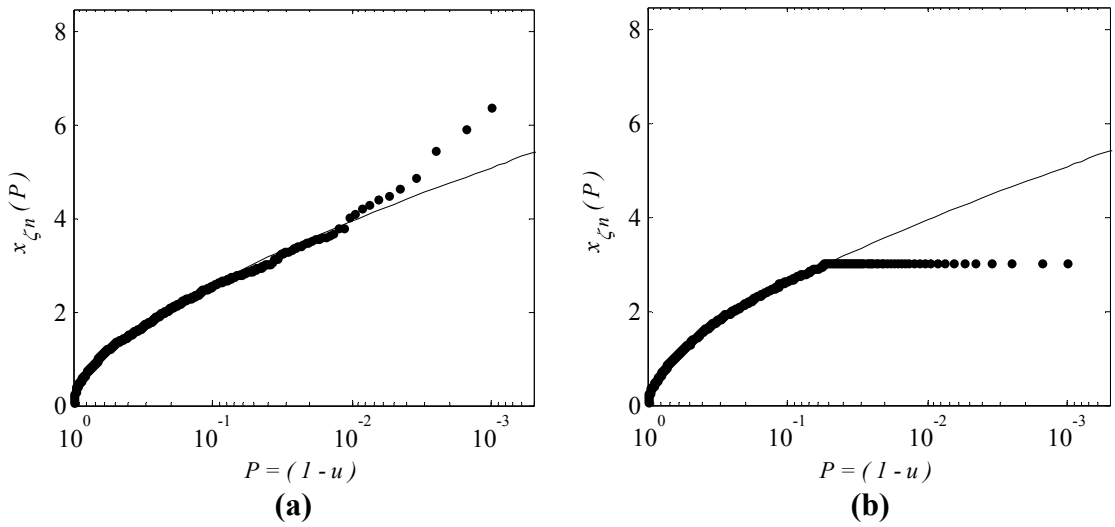


Figure 17 Samples with contaminated tails, (a) sample with unexpectedly large values, (b) sample with flat tail.

Twenty five thousands samples are generated from the distributions with faulty measurements utilizing Monte-Carlo simulation technique and the results for the samples with unexpectedly large observations and flat tails are respectively presented in Table 11 and Table 12. In these tables, the statistics of the first three moments and L-moments, estimated parameters with MoM and MoL, and expected maxima $E[\zeta_{\max}]$ are given. It should be noted that the uncertainty of these estimates is not only from the tail contamination but also from the sample variability. In that, the values in these tables should be compared with the results of sample variability analysis performed earlier.

The results obtained for these examples indicate that higher moments, i.e. μ_2 and μ_3, are more sensitive to the errors on the distribution tail as compared to the corresponding L-moments. This can be realized from comparing the statistics given in Table 11 and Table 12 with the ones presented earlier in Table 4 for variability of the

moments and L-moments of RS model. As shown here, the third moment is significantly biased and has considerably high variability which affects the parameter estimates of MoM. Comparing the results of these examples with the ones in Table 7, one can see that MoL is the more efficient parameter estimation method in estimating the parameters of samples with contaminated tails. As expected, the statistics of extreme values, e.g. expected maxima, are significantly affected by the tail part of the distributions. For both examples, the extreme estimates of MoL are found to be a better representative of the true values. Our studies indicate that one may obtain better results by post-processing the data sets and trying to remove (or replace) the faulty measurements. However, as mentioned earlier, in most cases recognizing the faulty measurements from actual extreme observations is a complex procedure.

Table 11 Effect of unexpectedly large values on the Rayleigh-Stokes model estimates.

Method of L-moments						Method of moments					
Estimate	True Value	σ	Δ	RMSE	RMSE %	Estimate	True Value	σ	Δ	RMSE	RMSE %
l_1	1.45	0.03	0.01	0.03	1.94	$\hat{\mu}_1$	1.45	0.03	0.01	0.03	1.94
l_2	0.47	0.01	0.01	0.01	3.03	$\hat{\mu}_2$	0.72	0.04	0.04	0.06	8.22
l_3	0.08	0.01	0.01	0.01	13.72	$\hat{\mu}_3$	0.56	0.12	0.16	0.20	35.26
$\hat{\alpha}$	1.00	0.07	-0.06	0.09	9.46	$\hat{\alpha}$	1.00	0.11	-0.13	0.17	17.06
$\hat{\beta}$	0.10	0.03	0.03	0.04	41.98	$\hat{\beta}$	0.10	0.04	0.06	0.07	69.27
$\hat{\gamma}$	0.00	0.04	0.02	0.04	NA	$\hat{\gamma}$	0.00	0.06	0.06	0.08	NA
$\hat{E}[\zeta_{\max}]$	5.37	0.23	0.23	0.33	6.13	$\hat{E}[\zeta_{\max}]$	5.37	0.27	0.39	0.48	8.88

Table 12 Effect of flat tail on the Rayleigh-Stokes model estimates.

Method of L-moments						Method of moments					
Estimate	True Value	σ	Δ	RMSE	RMSE %	Estimate	True Value	σ	Δ	RMSE	RMSE %
l_1	1.45	0.02	-0.03	0.04	2.55	$\hat{\mu}_1$	1.45	0.02	-0.03	0.04	2.55
l_2	0.47	0.01	-0.03	0.03	6.11	$\hat{\mu}_2$	0.72	0.02	-0.12	0.12	16.87
l_3	0.08	0.00	-0.03	0.03	35.03	$\hat{\mu}_3$	0.56	0.02	-0.35	0.35	61.90
$\hat{\alpha}$	1.00	0.07	0.21	0.22	21.96	$\hat{\alpha}$	1.00	0.06	0.35	0.35	35.07
$\hat{\beta}$	0.10	0.02	-0.10	0.11	106.21	$\hat{\beta}$	0.10	0.02	-0.16	0.16	157.17
$\hat{\gamma}$	0.00	0.03	-0.08	0.09	NA	$\hat{\gamma}$	0.00	0.03	-0.15	0.15	NA
$\hat{E}[\zeta_{\max}]$	5.37	0.13	-0.83	0.84	15.61	$\hat{E}[\zeta_{\max}]$	5.37	0.08	-1.15	1.15	21.46

CHAPTER IV

SEMI-EMPIRICAL MODEL APPLICATIONS

IV.1 Introduction

The semi-empirical models developed in this study are applicable to any weakly non-linear random variable that can be approximated by the second-order perturbation expansion Eq. (2.3). This covers a wide range of random variables in the field of ocean engineering and physical oceanography. The weakly non-linear surface elevation in deep and intermediate water depth and waves in the vicinity of structures are classic examples of random variables that can be approximated by the second-order perturbation expansion. In this chapter, the three semi-empirical models, i.e. Gaussian-Stokes (GS), Rayleigh-Stokes (RS), and Weibull-Stokes (WS) models, are applied to estimate the probability distribution of four weakly non-linear random variables. Particularly, the GS model is used to estimate the probability distribution of non-linear surface elevations η of undisturbed ocean waves, as well as disturbed waves in the vicinity of and underneath of offshore structures. Additionally, the RS and WS models are utilized to estimate the probability distribution of wave crests, a, of disturbed and undisturbed wave crests and wave run-up over the vertical columns of offshore structures, r. Both examples are studied for extremely energetic seastates, as well as, more benign seastates. In the final example, a special form of the RS model with no linear term ($\alpha = 0$) is utilized to estimate the probability distribution of wave power in benign environment. The performance of the semi-empirical models in capturing the probability distribution of

data is compared with appropriate theoretical and empirical models. The sample data sets utilized here are mainly obtained from a mini-TLP model test program investigating the response of the structure in random incident waves; when necessary, the samples are generated numerically.

In order to evaluate the performance of the semi-empirical and empirical models qualitatively, the uncertainty of the model statistics is estimated utilizing bootstrap analysis. Bootstrap analysis is a powerful computer-intensive method typically used to determine the bias and variance associated with a sample estimate ϕ_s of a parameter of interest ϕ [39, 40]. In this method independent samples of equal size to the original sample are generated from a distribution model with replacement assuming that observations are from an independent and identically distributed population. Here, the distribution model is approximated by a smoothed empirical distribution estimated from applying the non-parametric kernel density estimation method [46]. The results of different research studies comparing the properties of smoothed and non-smoothed bootstrap methods indicate that in most cases smoothing improves the bootstrap estimations and convergence [47- 49].

The extreme quantile estimation essentially needs a prediction of the distribution beyond the data set. In such cases, the standard non-parametric bootstrap method is not considered suitable since the non-parametric distribution model cannot provide a good approximation of the tail distribution. One effective way to overcome this problem has been proposed by Caers and Maes [41] in which the empirical tail is replaced by a parametric tail estimated from fitting a Generalized Pareto Distribution (GPD) to the

extreme values over a given threshold. The tail model, however, is sensitive to the selected threshold and requires iterative optimization. For this purpose, the procedure proposed by Caers and Maes [41] is followed in which the error in extreme quantile estimation is minimized. The formulation of the semi-parametric smoothed bootstrap method is briefly presented in Appendix A.

IV.2 Mini-TLP Model Test Data

The data sets utilized in this research study were mostly obtained from a model test investigating the response behavior of a mini-TLP in random seas. Selected particulars of the mini-TLP model are presented in Table 13 and additional details can be found in the articles [22, 50, 51]. The model tests were performed at a model scale of 1:40 and the prototype water depth was 668m. In Figure 18 the mini-TLP experimental setup is presented. The mini-TLP was originally designed for an environment off the coast of West Africa and had a prototype deck clearance elevation of 10m measured from mean water level. However, in order to study the behavior of the structure subject to the design sea conditions in the Gulf of Mexico the deck elevation was increased by 5m prototype scale for the model tests.

Table 13 Main particulars of the prototype mini-TLP.

Draft	28.50 m
Column Diameter	8.75 m
Column Spacing	28.50 m
Pontoon Height	6.25 m
Pontoon Width	6.25 m
Water Depth	668 m

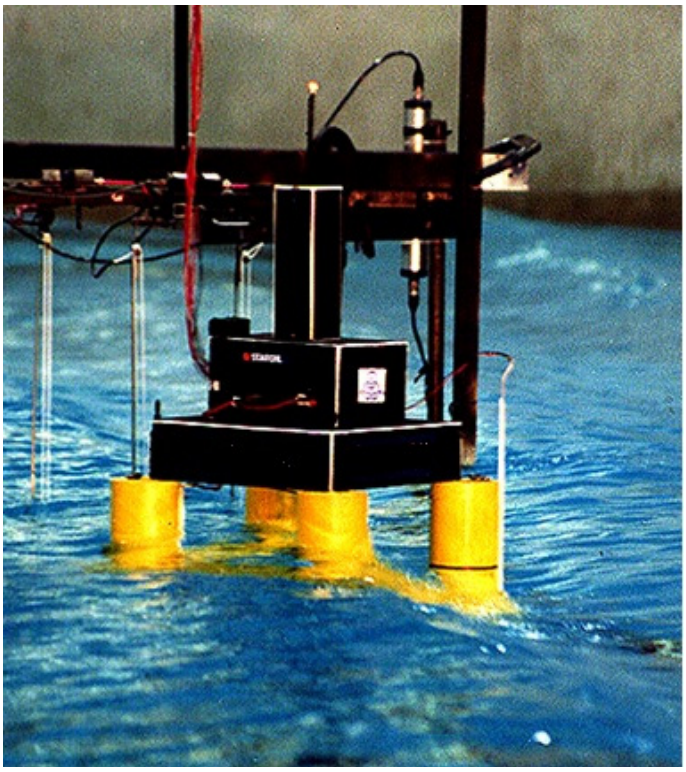


Figure 18 The mini-TLP experiment setup (courtesy of OTRC).

The 3-hr design seas were generated using a JONSWAP wave spectrum specified as a function of significant wave height H_s, peak period T_p, and spectral peakedness factor γ_s. For the first design storm where the sea was relatively benign the spectral parameters were specified as $H_s = 4m$, $T_p = 16s$ and $\gamma_s = 2$ (called S-1 hereafter) and for the second more energetic design storm the spectrum parameters were selected to be $H_s = 13.1m$, $T_p = 14s$ and $\gamma_s = 2.2$ (called S-2 hereafter).

The experimental study was conducted in two phases. Initially the motion response of a fully compliant TLP model, consisting of the hull, tendons and risers was instrumented and tested. During the second phase the compliant TLP model was disassembled and only the hull was rigidly suspended via a six-degree-of-freedom load cell assembly from the access bridge which itself was stiffened to prevent any movement. For the compliant model tests, the motion response of the system including the tendon and riser tensions, and the wave run-up were measured. For the fixed hull configuration the recorded time series included the applied forces and moments and wave run-up. Both configurations were subjected to identical regular, random unidirectional and directional seas. A schematic indicating the location of the reference wave probes, air-gap probes, and three surface adhered wave run-up probes used in the mini-TLP experiments is presented in Figure 19.

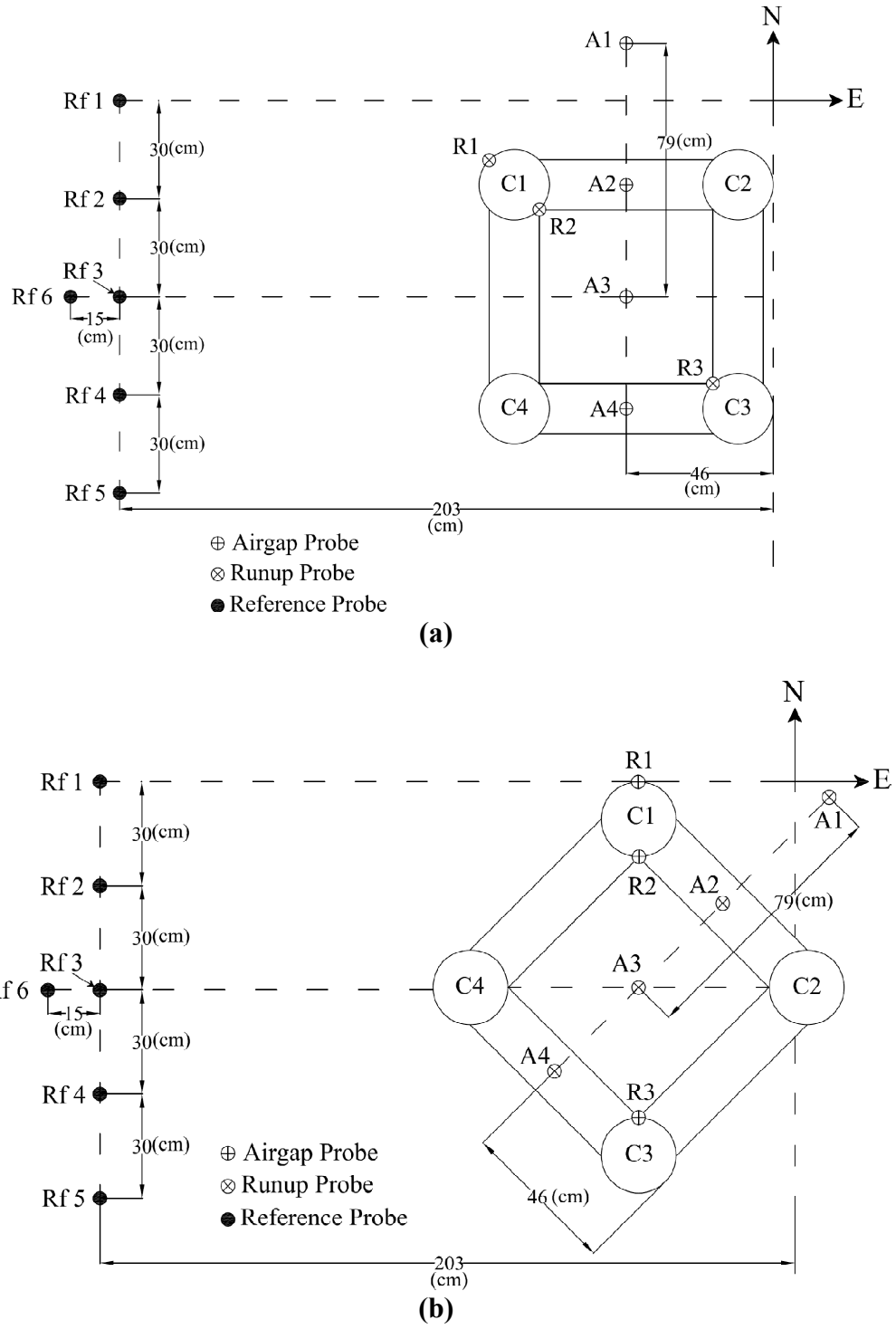


Figure 19 Plan view of the mini-TLP model test and the wave probe locations, (a) head sea, (b) quartering sea. Incident waves are propagating from North to South.

IV.3 Probability Distribution of Wave Elevations

IV.3.1 Introduction

The wave kinematics of the waves with low steepness can be adequately approximated by the linear wave theory, and that the linear surface elevation is nearly Gaussian. However, the probability distribution of surface elevation of steep waves deviates from Gaussian distribution due to the non-linear effects. Non-linear waves have sharper peaks and flatter troughs as compared to the linear waves and consequently their probability distributions are skewed towards positive values. The contribution of non-linear effects becomes more sensible in waves interacting with a structure. The intact incident waves elevate in the vicinity and underneath of an offshore/coastal structure due to the non-linear interaction between the incident waves and the diffracted and radiated waves from the structure. Engineers tasked to design the offshore structures in extreme environment need to have accurate estimates of the probability distribution of non-linear wave elevations in the vicinity of the structure.

In the literature there are numerous research studies on the theoretical probability distribution of non-linear wave elevations; however, a universally accepted theoretical probability distribution for non-linear wave elevation is not available. Among the earliest theoretical models, the Gram-Charlier probability distribution derived by Longuet-Higgins [7] is presumably one of the most well recognized models. The model is developed utilizing the cumulant generating function and application of the polynomials orthogonal with respect to the Gaussian probability distribution, i.e. Hermite polynomials. The Gram-Charlier probability distribution can be considered as a

moment-based model as the model parameters are directly related to the moments of the random process. In the original model the moments are related to the wave spectrum, however, the model can be used as an empirical probability distribution by replacing the moments with their corresponding sample estimates. The performance of a four-term Gram-Charlier probability distribution has been tested on different measured and numerically simulated data sets [7, 8, 52] and in general a good agreement has been observed. The main drawbacks of the Gram-Charlier model are that the model PDF becomes negative for large negative values and the original model estimates are biased towards negative values [52].

Another family of non-linear wave elevation probability distributions is developed based on the Stokes theory of small amplitude waves. Tayfun [8] developed the probability distribution of second-order modulated Stokes waves assuming that the linear waves have narrow-band spectrum. It was observed that the probability distribution fits well with simulated data sets [8] and experimental data sets [53]. The probability distribution, however, does not have a closed analytical form. Huang et al. [9] utilized a similar though different approach to obtain the probability distribution of third-order Stokes wave elevations assuming that waves are narrow-banded random process. The model has shown to have reasonable accuracy in modeling the probability distribution of non-linear wave elevations.

Another set of theoretical models was developed based on the concept of non-linear systems and approximating the system's response by a two-term Volterra stochastic series expansion [10, 54]. The model coefficients are estimated employing

Kac-Siegert analysis. The model is fairly robust while requiring a significant amount of calculations and the closed analytical form of the probability distribution is not available. Another family of models includes the distributions developed based on the Hermite transformation of the linear process [55-57]. The Hermite based models have been used to estimate the probability distribution of different random variables e.g. wave forces and structural motions, and the statistics found to be fairly accurate.

All the above mentioned theoretical models are derived for undisturbed waves and do not consider the effects of wave-structure interaction on the probability distribution of wave surface elevation. The mathematical background and wave theories, however, are still valid for waves interacting with a structure. The difficulty arises in estimating the model parameters. Almost all of the theoretical models utilize the wave energy spectrum or wave number spectrum to estimate the model parameters. In case of intact waves, the wave spectrum is commonly approximated by a parametric wave spectrum, e.g. JONSWAP, Pierson-Moskowitz, etc., that is a function of a few seastate characteristics, e.g. significant wave height, peak period, etc. [58]. Such parametric spectrums are not available for waves in the vicinity of structures mainly because of the variety in the structural type, shape, size, and configuration and the complexity of the wave field around the structure. In that, the theoretical models in their original form cannot be applied to model the probability distribution of wave elevations in the vicinity and underneath the structure and designers routinely rely on empirical models for this purpose.

Gaussian distribution has been widely used for engineering purposes in which the distribution parameters can be easily estimated from the sample mean and standard deviation. The main problem with Gaussian probability distribution is that it cannot model the skewness and kurtosis in the non-linear waves. A more appropriate option for weakly non-linear waves is the beta distribution introduced by Srokosz [59]. The beta distribution models the skewness in the data while assuming that the excess kurtosis is zero. Although the model parameter of the beta distribution was originally estimated utilizing the theoretical wave elevation skewness derived by Longuet-Higgins [7], it can be empirically estimated from application of the sample skewness.

As an alternative to the theoretical and empirical models, the Gaussian-Stokes (GS) semi-empirical model is utilized in this study to estimate the probability distribution of wave elevations of non-linear ocean waves as well as waves in the area around an offshore structure. The GS model can be classified in the family of two-term Volterra models in which the dynamic system response is approximated by summation of the linear and quadratic terms. The GS model parameters are then estimated empirically utilizing the sample statistics obtained from wave elevation timeseries. In the followings the mathematical background of the Stokes wave theory and theoretical and empirical probability distributions are briefly overviewed and then the application of the GS model for simulated and measured wave data is discussed.

IV.3.2 Mathematical background

In the theory of linear waves, the surface elevation of unidirectional irregular waves in deepwater is represented by Fourier series as [58],

$$\eta_1(x,t) = \mathrm{Re}\left(\sum_n^{\infty} c_n e^{i\chi_n} \right) \tag{4.1}$$

where c_n are the Fourier constants, $\chi_n = k_n x - \omega_n t + \varepsilon_n$ are the total phases, ω_n are the angular frequencies, $k_n = 2\pi/\lambda_n$ are wave numbers, λ_n are wave lengths, and ε_n represent random phases uniformly distributed over the interval of $0 - 2\pi$. The relation between the wave numbers k_n and frequencies in deepwater is defined by dispersion relation $\omega_n^{\,2} = g k_n$. This representation of surface elevation essentially assumes that the first-order process η_1 is a Gaussian stochastic process with zero mean and standard deviation of σ_η. A second-order solution of the wave elevation η may be obtained from perturbation approximation of the small-amplitude-waves boundary value problem [42]. The perturbation solution is represented as the summation of the first- and second-order terms, specifically,

$$\eta(x,t) = \eta_1(x,t) + \eta_2(x,t) \tag{4.2}$$

where the second-order term $\eta_2(x,t)$ is represented in the form of,

$$\eta_2(x,t) = \mathrm{Re}\left(\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} c_n c_m \left[r_{nm} e^{i(\chi_n + \chi_m)} + q_{nm} e^{i(\chi_n - \chi_m)} \right] \right) \tag{4.3}$$

The equations for the parameters r_{nm} and q_{nm} for general water depth can be found in Appendix A of the article [10]. The asymptotic relations for these parameters in deepwater are simplified in the form of,

$$\begin{aligned} r_{nm} &= \left(\frac{1}{4g}\right)\left(\omega_n{}^2 + \omega_m{}^2\right) \\ q_{nm} &= -\left(\frac{1}{4g}\right)\left(\omega_n{}^2 - \omega_m{}^2\right)\mathrm{sgn}\left(\omega_n - \omega_m\right) \end{aligned} \tag{4.4}$$

where $\mathrm{sgn}(x)$ denotes the sign of x. As shown in Eq. (4.3) the second-order term consists of a high frequency term oscillating at the sum frequencies of the linear terms, and a low frequency term oscillating at the difference frequencies.

The spectral moments are routinely used to estimate the wave characteristics in the frequency domain. The jth order spectral moment can be estimated from,

$$m_j = \int_0^\infty \omega^j G_{\eta\eta}(\omega)\, d\omega \tag{4.5}$$

where $G_{\eta\eta}(\omega)$ is the single-sided wave frequency spectrum accurate to the first-order, and is obtained from,

$$\sum_{\omega \le \omega_n < \omega + d\omega} c_n{}^2 \approx 2G_{\eta\eta}(\omega)\, d\omega \tag{4.6}$$

The contribution of the second-order terms in $G_{\eta\eta}$ is proportional to g^{-2} and is negligible. The first-order variance of η and the spectral mean frequency are among useful statistics that can be defined in terms of the spectral moments $\sigma_\eta{}^2 = m_0$, and

$\bar{\omega} = m_1 / m_0$, respectively. A measure for the spectral width ν_s is proposed by Longuet-Higgins [60, 61] which cooperates the first three spectral moments, specifically,

$$\nu_s^{\,2} = \left(\frac{m_2 m_0}{m_1^{\,2}} - 1 \right) \tag{4.7}$$

A wave spectrum is considered to be narrow-banded when $\nu_s^{\,2} \ll 1$. The narrow-banded approximation has been widely used to model the random processes in the field of ocean engineering. For a narrow-banded random process, the first-order wave elevation $\eta_1(x,t)$, Eq. (4.1) , can be approximated by,

$$\eta_1(x,t) = a(x,t)\cos\bar{\chi} \tag{4.8}$$

where $a(x,t)$ and $\bar{\chi}$ are amplitude and phase function, respectively. It can be shown that the narrow-band representation of second-order term $\eta_2(x,t)$, Eq. (4.3), is in the form of [8],

$$\eta_2(x,t) = \frac{1}{2}\bar{k}\,a^2(x,t)\cos 2\bar{\chi} \tag{4.9}$$

where $\bar{k}$ is the wave number corresponding to the mean frequency.

The non-linear waves at a point can also be considered as the response of a non-linear dynamic system to a Gaussian input. The response of the non-linear dynamic system has been routinely approximated by Volterra stochastic series expansion. Kac and Siegert [62] analytically showed that the two-term Volterra stochastic series expansion can be represented as a sum of statistically independent random variables Z_i, specifically,

$$\eta(t) = \sum_{i=1}^{N} Z_i = \sum_{i=1}^{N} \left(\Upsilon_i z_i + \Lambda_i z_i^2 \right) \tag{4.10}$$

where z_i are uncorrelated Gaussian random variables with zero mean and unit variance. Langley [10] evaluated the parameters Υ_i and Λ_i utilizing the second-order Stokes wave theory and the concept of wave-wave interaction. For this purpose he presented the second-order wave elevation, Eq. (4.2), in a matrix form and then applied eigenvalue analysis to obtain the functional relationship between the parameters Υ_i and Λ_i and the interaction coefficients. A simplified form of this representation can be obtained from utilizing a single variable instead of summation of the standardized Gaussian variables, as,

$$\eta(t) = \alpha z + \beta z^2 \tag{4.11}$$

where, again z is a Gaussian random variable with zero mean and unit variance. The model is identical to the perturbation expansion used to develop the GS semi-empirical model, Eq. (2.3) with $\gamma = 0$.

IV.3.3 Theoretical probability distributions

The performance of the GS model in capturing the probability distribution of non-linear ocean wave elevation is compared with three well known theoretical probability distributions of non-linear wave elevations, i.e. Gram-Charlier distribution developed by Longuet-Higgins [7], second order narrow-band model [8], and third-order narrow-band model [9]. These models are briefly overviewed here and for more details one is referred to the original papers.

Longuet-Higgins [7] approximated the probability distribution of normalized non-linear wave elevation $\zeta_n = \eta/\sigma_\eta$ in terms of the Gram-Charlier series, specifically,

$$f_{\zeta_n}(x) = \frac{1}{(2\pi)^{1/2}} e^{-x^2/2}\left[1 + \frac{1}{6}s_\eta H_3(x) + \frac{1}{24}K_\eta H_4(x) + \frac{1}{72}s_\eta{}^2 H_6(x) + \cdots\right] \tag{4.12}$$

where $s_\eta = \mu_3(\eta)/\sigma_\eta^3$ and $K_\eta = \mu_4(\eta)/\sigma_\eta^4$ are the wave elevation skewness and kurtosis, respectively, and H_n denotes the Hermite polynomial of degree n, where for $n =$ 3, 4 and 6 are defined as,

$$\begin{aligned} H_3(x) &= x^3 - 3x, \\ H_4(x) &= x^4 - 6x^2 + 3, \\ H_6(x) &= x^6 - 15x^4 + 45x^2 - 15. \end{aligned} \tag{4.13}$$

The theoretical estimates of the wave elevation third and fourth order moments, i.e. $\mu_3(\eta)$ and $\mu_4(\eta)$ are obtained from [7, 63, 64]

$$\begin{aligned} \mu_3(\eta) &= \frac{3}{g}\sum_{n=1}^{N_f}\sum_{m=1}^{N_f}\sigma_n{}^2\sigma_m{}^2 \min\left(\omega_n^2, \omega_m^2\right), \\ \mu_4(\eta) &= \frac{12}{g^2}\sum_{n=1}^{N_f}\sum_{m=1}^{N_f}\sum_{k=1}^{N_f}\sigma_n{}^2\sigma_m{}^2\sigma_k{}^2 \min\left(\omega_n^2, \omega_k^2\right)\min\left(\omega_m^2, \omega_k^2\right). \end{aligned} \tag{4.14}$$

where N_f is the total number of frequencies and $\sigma_n = \left(G_{\eta\eta}(\omega_n)\Delta\omega_n\right)^{1/2}$. In case of narrow-banded waves in deepwater the expressions given in Eq. (4.14) can be approximated by [8],

$$\begin{aligned} \mu_3(\eta) &\approx 3\bar{k}m_0{}^2, \\ \mu_4(\eta) &\approx 3m_0{}^3\left(6\bar{k}^2 + 3\bar{k}^4 m_0\right) \end{aligned} \tag{4.15}$$

Different research studies indicate that the Gram-Charlier series up to the fourth term (i.e. $1/72\, s_\eta^{\,2} H_6(x)$) is the optimum approximation [8, 9].

The second theoretical probability distribution that is studied here is the probability distribution derived by Tayfun [8] based on the assumptions that waves have narrow-band frequency spectrum and can be approximated by second-order Stokes expansion Eq. (4.8) and (4.9). In this model the zero-mean normalized wave elevation is represented by,

$$\zeta_n = z_1 + \frac{1}{2}\bar{k}\, m_0^{1/2}\left(z_1^{\,2} - z_2^{\,2}\right) \tag{4.16}$$

where z_1 and z_2 are zero-mean Gaussian random independent variables with the joint PDF of,

$$f_{z_1,z_2}(x_1,x_2) = \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{1}{2}\left(x_1^{\,2} + x_2^{\,2}\right)\right) \tag{4.17}$$

The CDF of ζ_n is then estimated by numerical evaluation of,

$$F_{\zeta_n}(x) = \text{Prob}\left(z_1 + \frac{1}{2}\bar{k}\, m_0^{1/2}\left(z_1^{\,2} - z_2^{\,2}\right) \le x\right) \tag{4.18}$$

The probability distribution in Tayfun's model does not have a closed form representation.

Huang et al. [9] derived the closed form probability distribution of the third-order narrow-banded Stokes waves in deepwater. They developed two models: "Stokes wave in deep water with bias" and "Stokes wave in deep water without bias". The term bias in these models refers to the constant term $1/2\,\bar{k}\,a^2$ in the third-order Stokes expansion.

Huang et al. [9] indicated that the constant term has important effects on the probability distribution of non-linear wave elevations. However, the results of other research studies questioned the validity of the model with the constant term, i.e. model with bias, and indicated that the model without bias is the more robust alternative [63, 65]. The PDF of the non-linear normalized wave elevation in the Huang et al model without the bias term is given as,

$$f_{\zeta_n}(x)=\frac{\exp(-H_x)}{(2\pi)^{1/2}}\left[J_x\left(\frac{1}{R_x^{1/2}}-\frac{3}{8}\frac{m_0\bar{k}^2}{N_x^2}\frac{1}{R_x^{5/2}}\right)+\frac{5}{8}\frac{m_0\bar{k}^2}{N_x^2}\frac{1}{R_x^{3/2}}\right] \tag{4.19}$$

where,

$$\begin{aligned}
N_x &= 1+\frac{1}{2}m_0\bar{k}^2,\\
R_x &= 1+m_0^{1/2}\bar{k}x+\frac{3}{4}m_0\bar{k}^2x^2,\\
J_x &= N_x\left(1-m_0^{1/2}\bar{k}x+\frac{3}{8}m_0\bar{k}^2x^2\right),\\
H_x &= N_x^2\left(x^2-m_0^{1/2}\bar{k}x^3+\frac{3}{8}m_0\bar{k}^2x^4\right).
\end{aligned} \tag{4.20}$$

IV.3.4 Empirical probability distributions

Gaussian probability distribution has been widely used to estimate the probability distribution of wave elevation records. However, Gaussian distribution cannot correctly model the skewness and kurtosis of the non-linear elevation. This limits the application of empirical Gaussian distribution to low steepness waves. The PDF of $\zeta_n=\eta/\sigma_\eta$ approximated by Gaussian probability distribution is defined as,

$$f_{\zeta_n}(x) = \frac{1}{(2\pi)^{1/2}\,\sigma_{\zeta_n}} \exp\left(-\frac{x^2}{2\sigma_{\zeta_n}{}^2}\right) \tag{4.21}$$

Note that here the wave elevation data is normalized by the first-order standard deviation σ_η to be consistent with the theoretical models. Therefore, the normalized elevation ζ_n has a zero mean and a standard deviation of σ_{ζ_n} which is not essentially equal to one due to the contribution of the non-linear terms. The empirical estimates of σ_{ζ_n} can be obtained from its relation with the sample second moment $\hat{\mu}_2(\zeta_n)$ and second L-moment $l_2(\zeta_n)$, specifically,

$$\begin{aligned} \hat{\sigma}_{\zeta_n}{}^2 &= \hat{\mu}_2(\zeta_n), \\ \hat{\sigma}_{\zeta_n} &= \pi^{1/2} l_2(\zeta_n). \end{aligned} \tag{4.22}$$

The next empirical probability distribution that is studied here is the beta probability distribution developed by Srokosz [59] for weakly non-linear wave elevations. Srokosz [59] utilized the Pearson system of distributions to obtain a particular form of the generalized beta distribution with non-zero skewness and zero kurtosis. Based on these assumptions the PDF of $\zeta_n = \eta/\sigma_\eta$ is given as,

$$f_{\zeta_n}(x) = \begin{cases} \dfrac{\Gamma(q'+q)}{\Gamma(q')\Gamma(q)} \dfrac{(x-\gamma_1)^{q'-1}(\gamma_2-x)^{q-1}}{(\gamma_2-\gamma_1)^{q'+q-1}} & \gamma_1 \le x < \gamma_2 \\ 0 & x < \gamma_1 \; and \; x > \gamma_2 \end{cases} \tag{4.23}$$

where γ_1 and γ_2 denote the lower and upper limits of the distribution, respectively, and both parameters q' and q are the shape parameters. The relation between γ_1 and γ_2 and the normalized wave elevation skewness s_{ζ_n} is given as,

$$\gamma_1 = \frac{1}{s_{\zeta_n}}\left(1-\left(5-{s_{\zeta_n}}^2\right)\right),$$
$$\gamma_2 = \frac{1}{s_{\zeta_n}}\left(1+\left(5-{s_{\zeta_n}}^2\right)\right). \tag{4.24}$$

and the shape parameters are estimated from,

$$q' = 1+\frac{\left(\gamma_0+\gamma_1\right)}{\gamma_{c_2}\left(\gamma_2-\gamma_1\right)},$$
$$q = 1-\frac{\left(\gamma_0+\gamma_2\right)}{\gamma_{c_2}\left(\gamma_2-\gamma_1\right)}, \tag{4.25}$$

where,

$$\gamma_0 = \frac{s_{\zeta_n}}{2\left(1-{s_{\zeta_n}}^2\right)}, \qquad \gamma_{c_2} = \frac{-{s_{\zeta_n}}^2}{4\left(1-{s_{\zeta_n}}^2\right)} \tag{4.26}$$

The model requires the empirical estimate of s_{ζ_n} which can be obtained from its relation with the sample's second and third moments as,

$$\hat{s}_{\zeta_n} = \frac{\hat{\mu}_3\left(\zeta_n\right)}{\left(\hat{\mu}_2\left(\zeta_n\right)\right)^{3/2}} \tag{4.27}$$

Note that the relations in Eq. (4.26) are singular at $\hat{s}_{\zeta_n} = 1.0$ and consequently, the beta distribution is valid for $0 \le \hat{s}_{\zeta_n} < 1$. More importantly, the PDF of wave elevations is expected to have a "humped-shape" and the PDF of beta distribution is only humped-shaped for $q' > 1$ and $q > 1$. This limits the application of beta model for wave records with $0 \le \hat{s}_{\zeta_n} < 0.8265$ [59]. The Srokosz beta distribution asymptotes to Gaussian distribution for $\hat{s}_{\zeta_n} \to 0$. The results of our analysis indicate that for $0 \le \hat{s}_{\zeta_n} < 0.1$ the statistics of beta and Gaussian distributions are almost identical.

IV.3.5 Sample data sets

The data sets utilized here are mainly from the mini-TLP model experiment described in IV.2. Both undisturbed wave records measured at reference probe RF6 (see Figure 19), and disturbed waves measured at air gap probes A2-A4 are studied here. Additionally, the probability distribution of a numerically generated second-order wave record is studied. The second-order wave record is generated by Monte-Carlo simulation of the second-order wave interaction equation, Eq.(4.2), with uniformly distributed random phases. It is assumed that the linear waves have a JONSWAP power spectrum and the simulated waves represent a unidirectional wave field in deepwater. The simulated realization models a 3-hr seastate with a sampling frequency of $f_s = 5\,\text{Hz}$. The characteristics of the studied seastate including the significant wave height H_s, peak period T_p, JONSWAP spectrum peakedness factor γ_s, average period $\overline{T}$, and spectrum width ν_s^2 (Eq. (4.7)) are presented in Table 14. The first two seastates are the ones utilized in the mini-TLP model test and respectively represent the 100yr return period seastates in West Africa and Gulf of Mexico. The third spectrum is used in the second-order wave simulation and represents a highly energetic seastate with relatively narrow-banded spectrum.

Table 14 Basic characteristics of the studied seastates.

Statistics	Seastate		
	Model Test		Numerical Simulations
	S-1	S-2	S-3
H_s (m)	4.0	13.1	13.1
T_p (sec)	16.0	14.0	9.0
γ_s	2.2	2.0	7.0
$\overline{T}$ (sec)	13.0	11.3	7.9
ν_s^2	0.2	0.2	0.1

IV.3.6 Analysis and results

IV.3.6.1 Undisturbed waves

The performance of GS semi-empirical model is firstly evaluated over undisturbed waves measured at the reference probe Rf-6, see Figure 19, and the simulated second-order waves. The assumptions of the theoretical models are well justified in these cases and consequently the GS model statistics can be directly compared to those estimated from the theoretical models. The theoretical models used here are the standard Gaussian distribution of linear process (zero mean and unit standard deviation), Longuet-Higgins' Gram-Charlier (GC) model (Eq.(4.12)), Tayfun's second-order approximation (TY) (Eq. (4.18)), and Huang's third-order approximation (HU) (Eq. (4.19)). Additionally, the results of the empirical Gaussian (Eq. (4.21)) and beta distribution (Eq. (4.23)) for these examples are evaluated.

In Table 15, the numerical values of the first four L-moments and moments, of the normalized surface elevation $\zeta_n = \eta / \left(m_0^{1/2} \right)$ are presented. For each seastate, the first-order variance of surface elevation m_0 is estimate from its relationship with significant wave height $m_0 = H_s^{\,2} / 16$. It should be noted that the values given in Table 15 are estimated from a 3-hr realization (more than 50,000 sample points). The wave elevation sample measured at RF6 during the seastate S-1, the wave elevation sample measured at RF6 during the seastate S-2, and the numerically generated second-order wave record during the seastate S-3 are called S1W, S2W, and S3W, respectively. In Table 15 the statistics of the Gaussian distribution of the linear process are given for reference. The difference between the second moment (L-moment) of Gaussian process and the second moment of a wave record can be an indication of the contribution of the non-linear terms to the wave spectrum. As can be seen from the values in Table 15, the wave records S1W and S2W have the largest second moment, which may be caused by a combination of the higher order terms and the diffracted waves in the wave basin. The Gaussian distribution is symmetric around its mean and consequently has zero skewness (L-skewness). As mentioned earlier, non-linear terms cause sharper peaks and flatter troughs and consequently a skewed surface elevation distribution towards positive values. The wave sample S1W have the smallest skewness denoting that undisturbed waves at this sea-state are almost symmetrical. This is theoretically justified since waves in this seastate have the lowest steepness as compared to waves in the other two seastates. The other two wave records, i.e. S2W to S3W, are noticeably non-linear.

A positive kurtosis indicates that the PDF has a sharper peak as compared to the Gaussian PDF and more of the variance is caused by the extreme observations; the negative kurtosis indicates the opposite. In case of L-kurtosis a value larger or smaller than the Gaussian distribution L-kurtosis, i.e. 0.123, is equivalent to positive and negative kurtosis, respectively. In order to characterize the random variable's tail distribution the estimates of both skewness and kurtosis are required. As shown in Table 15, all the studied samples have positive kurtosis and the numerically simulated sample S3W has the largest kurtosis.

Table 15 L-moments and moments of normalized undisturbed surface elevation.

Statistics	Gaussian Process	Undisturbed wave elevation samples		
		S1W	S2W	S3W
$l_1(\zeta_n)$	0.000	0.000	0.000	0.000
$l_2(\zeta_n)$	0.564	0.597	0.595	0.584
$\hat{\tau}_3(\zeta_n)$	0.000	0.011	0.047	0.067
$\hat{\tau}_4(\zeta_n)$	0.123	0.128	0.124	0.129
$\hat{\mu}_1(\zeta_n)$	0.000	0.000	0.000	0.000
$\hat{\mu}_2(\zeta_n)$	1.000	1.127	1.116	1.088
$\hat{s}_{\zeta_n}$	0.000	0.068	0.288	0.398
$\hat{K}_{\zeta_n}$	0.000	0.119	0.099	0.316

Utilizing the values given in Table 15, the three parameters of the GS model are evaluated utilizing the method of moments (MoM) and method of L-moments (MoL) and the results for the three samples are presented in Table 16. As shown in this table,

the parameter estimates of MoM and MoL are almost identical for these samples. This is mainly because of the considerably large sample sizes and the smoothly distributed extreme values. The GS semi-empirical model has detected slightly larger linear contribution than 1.0 in the wave records. This can be realized from the estimates of the parameter α given in Table 16. Regarding the values estimated for parameter β, it can be understood that the third sample has the highest second-order contribution. The shifting parameter γ was found to be relatively small in all the three samples.

Table 16 Empirically estimated parameters of Gaussian-Stokes model (undisturbed wave samples).

Statistics	Method	Undisturbed wave elevation samples		
		S1W	S2W	S3W
$\hat{\alpha}$	MoL	1.059	1.054	1.036
	MoM	1.061	1.054	1.038
$\hat{\beta}$	MoL	0.012	0.051	0.071
	MoM	0.012	0.051	0.069
$\hat{\gamma}$	MoL	-0.012	-0.051	-0.071
	MoM	-0.012	-0.051	-0.069

In Figure 20 the exceedance probability distribution of normalized wave elevation of the sample S1W is presented. The focus here is on the positive tail of the distribution. The sample distribution in this figure (solid black line) is estimated utilizing the semi-parametric approach (see Appendix A) and the same distribution is applied in the bootstrap analysis. The dashed lines in Figure 20 denote the 95% confidence limits of the sample distribution estimated utilizing 10,000 bootstrap samples. The theoretical

distributions are presented in Figure 20 (a) and the GS semi-empirical and empirical distributions are shown in Figure 20 (b). As can be seen in Figure 20 (a), the distributions of theoretical models TY, GC, and HU are fairly close. The sample distribution in this case is almost symmetric about the mean value but has a heavier tail than the Gaussian distribution. A reasonable agreement between the theoretical models and the sample distribution is observed. However, the theoretical models slightly underestimate the sample tail distribution with exceedance probability smaller than 10^{-2}.

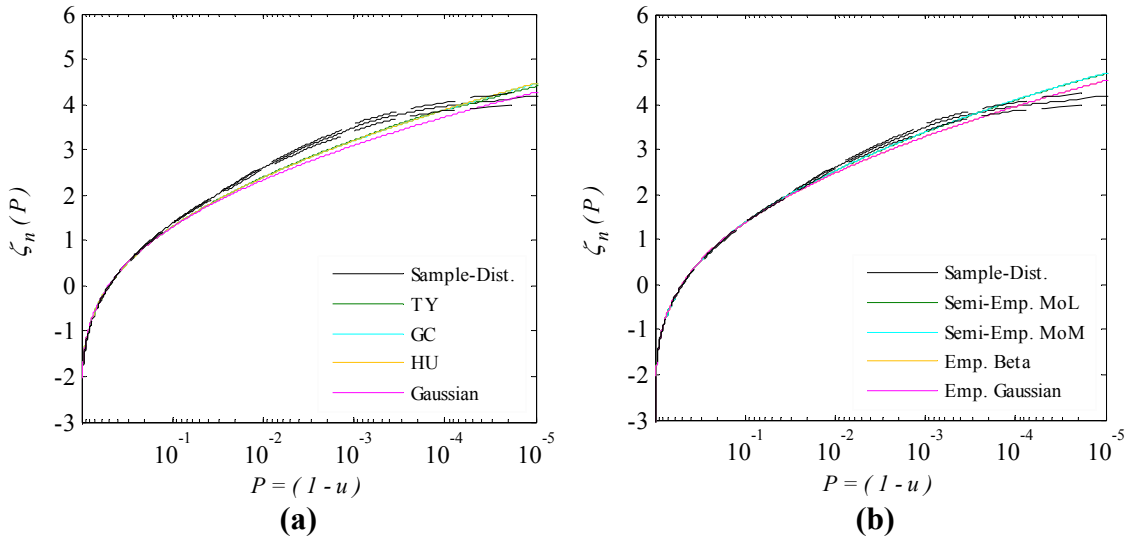


Figure 20 Exceedance probability distributions of the undisturbed wave sample S1W, (a) theoretical distributions, (b) semi-empirical and empirical distributions. Dashed lines represent the 95% confidence limits.

As shown in Figure 20 (b), the estimates of the GS semi-empirical distributions with the parameters estimated by MoL and MoM methods are almost identical. The GS semi-empirical model is robust in estimating the probability distribution of the weakly

non-linear wave elevation sample S1W. Although the three-parameter GS semi-empirical approach only models the first three moments (L-moments) of the wave record, it is fairly successful in modeling the sample tail distribution. Because of considerably small skewness of this wave record, the statistics of the empirical beta and Gaussian distributions are identical in this example. The empirical Gaussian distribution performs better than the standard Gaussian distribution of linear process as the empirical model considers the contribution of non-linear terms on the wave standard deviation.

The RMSE distributions of the quantile estimates for the semi-empirical GS and empirical beta models are presented in Figure 21. The RMSE values are estimated utilizing the 10,000 bootstrap samples. As shown in this figure, for most part of the distribution, the RMSE of the GS semi-empirical model remains below 0.1 (equivalent to 0.1m). As shown in Figure 20, the slope of sample distribution changes around $P = 9\times10^{-4}$ and the tail distribution becomes flatter. The semi-empirical distribution intersects with the sample distribution around $P = 3\times10^{-4}$ and overestimates the quantiles for smaller exceedance probabilities. The same pattern can be easily seen in the RSME distribution of the GS semi-empirical models given in Figure 21. For the most part of the distribution the RMSE of the GS model is smaller than the RMSE of the beta distribution while the beta distribution has smaller error for quantiles with $P < 2\times10^{-4}$.

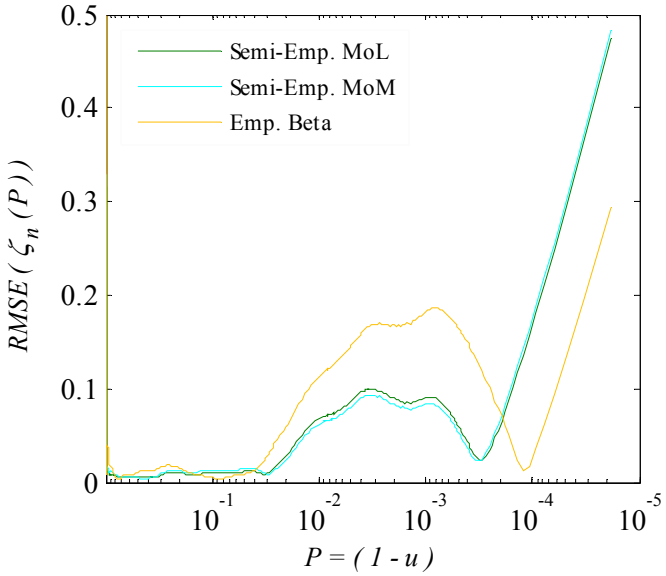


Figure 21 RMSE of the quantile estimates of the undisturbed wave sample S1W.

In Figure 22, similar to Figure 20, the exceedance probability distribution of the wave sample S2W is studied. The waves in this case are highly energetic and as indicated in Table 15, the observations are positively skewed and have a relatively small kurtosis. As shown in Figure 22, the theoretical models, i.e. TY, GC, and HU, and specially the Longuet-Higgins Gram-Charier distribution (GC) accurately represent the probability distribution of data. As expected, the standard Gaussian distribution underestimates the extreme observations. Once more, the semi-empirical models found to be robust in estimating the probability distribution of wave record and the statistics of GS model with the parameters estimated by MoL and MoM are almost identical. Regarding the distributions given in Figure 22 (b), one can easily see that the empirical beta distribution significantly underestimates the extreme observations.

The RMSE of the quantile estimates of the semi-empirical models and the beta model are presented in Figure 23. As shown here, the RMSE of the GS semi-empirical models remain below 0.05 (equivalent to 0.16m) for the most part of the distribution while the GS model overestimate the extreme values with $P < 3\times10^{-4}$. The estimates of beta distribution found to be unreliable in this case.

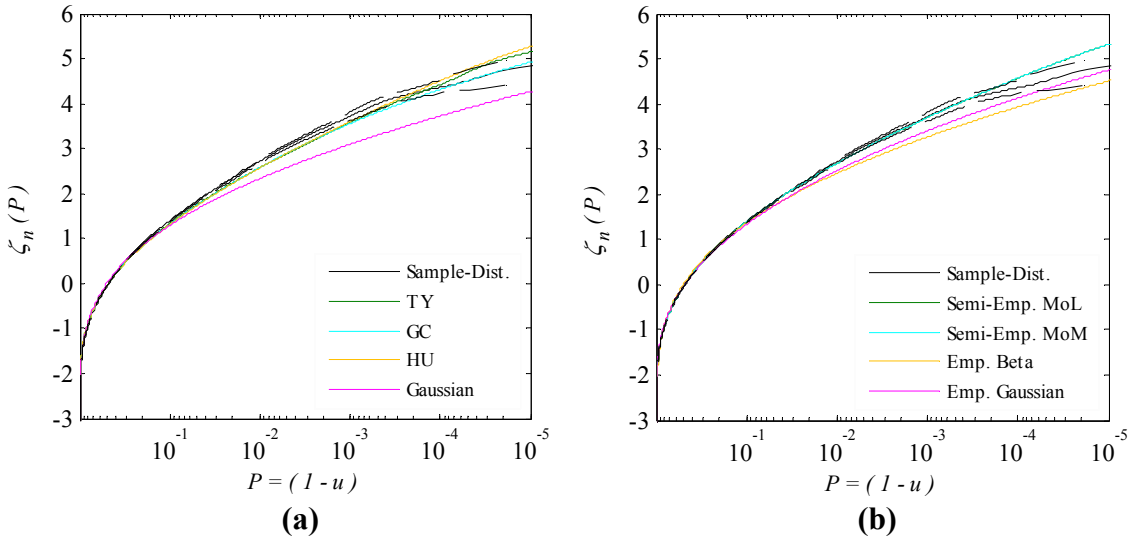


Figure 22 Exceedance probability distributions of the undisturbed wave sample S2W (a) theoretical distributions, (b) semi-empirical and empirical distributions. Dashed lines represent the 95% confidence limits.

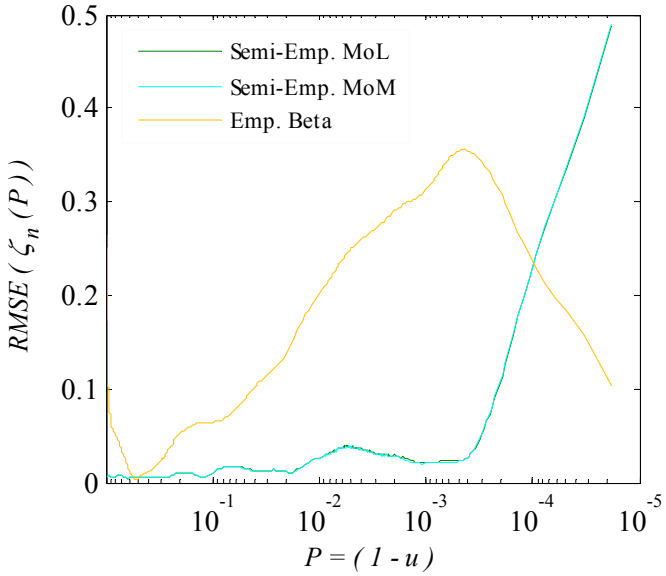


Figure 23 RMSE of the quantile estimates of the undisturbed wave sample S2W.

In the third example the probability distribution of simulated second-order wave elevation is studied. The waves in this case are highly energetic and considerably steep for deepwater condition. The exceedance probability distributions of the wave sample S3W are given in Figure 24. As can be seen in this figure, the non-linear theoretical models are robust in modeling the probability distribution of second-order waves. The GS semi-empirical models are also successful in capturing the probability distribution of data while Gaussian and beta distribution found to be unreliable especially in estimating the probability distribution of large wave elevations. Similar to Figure 21 and Figure 23, the RMSE of the quantile estimates of the semi-empirical and empirical beta models are presented in Figure 25 and a similar pattern has been observed. The semi-empirical models found to be reasonably accurate in modeling the overall probability distribution of data while they tend to overestimate the tails. As shown in Figure 24 for S3W as well

as Figure 22 for S2W, the sample distributions become flatter on the very end part of the distribution $P < 3\times10^{-4}$ which is not captured by the semi-empirical models. The sample flat tail may be caused by the increased contribution of the second-order negative terms (oscillating at difference frequencies) in extremely large wave elevations, which reduces the effect of positive terms. The contribution of the negative terms is not considered in the semi-empirical formulation which consequent the overestimation of the extremely large elevations.

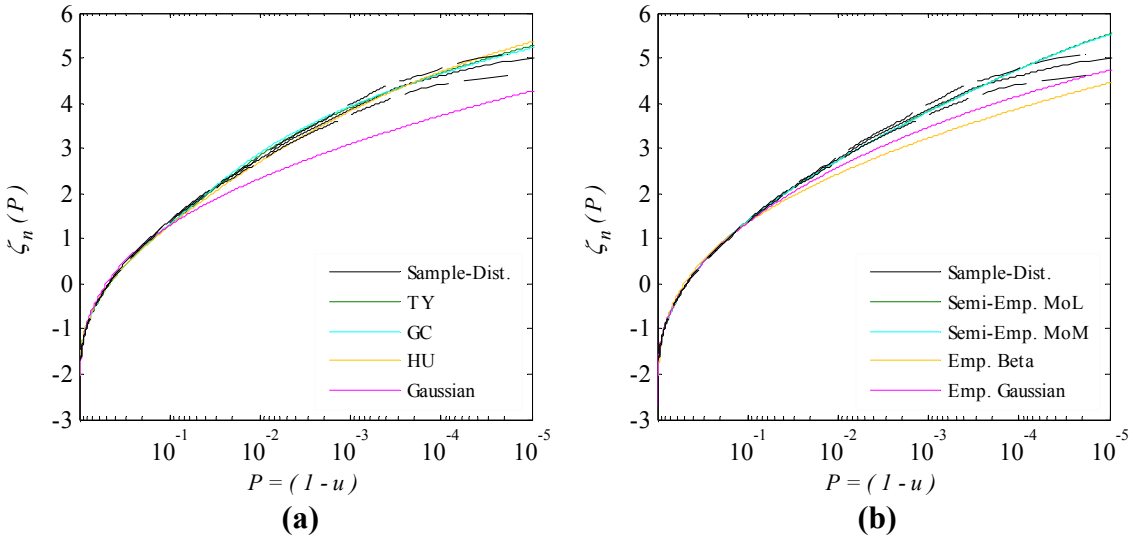


Figure 24 Exceedance probability distributions of the undisturbed wave sample S3W (a) theoretical distributions, (b) semi-empirical and empirical distributions. Dashed lines represent the 95% confidence limits.

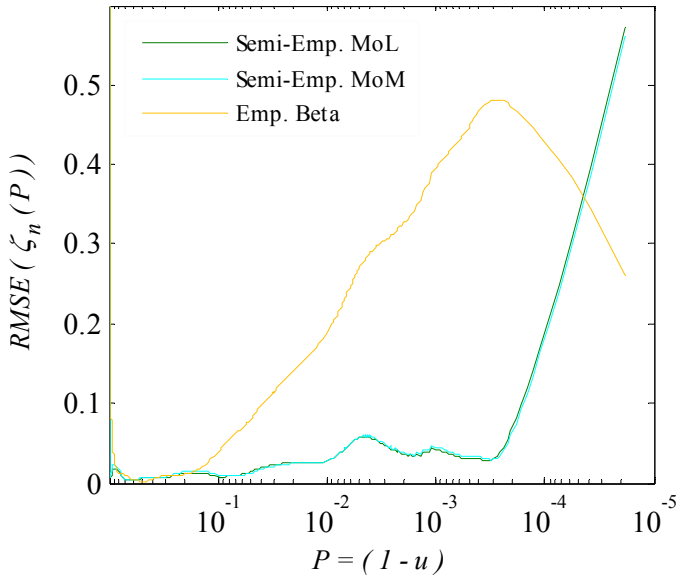


Figure 25 RMSE of the quantile estimates of the undisturbed wave sample S3W.

As shown in these three examples, the performance of the GS semi-empirical model on capturing the probability distribution of surface elevation of undisturbed waves is comparable to widely used theoretical models. Additionally, the GS model is more robust in capturing the probability distribution of data as compared with the widely used empirical distributions. The main purpose of this study is to estimate the probability distribution of the disturbed wave elevations interacting with offshore structures. The theoretical models in their initial formulation cannot be used for disturbed wave elevations providing that the models do not consider the effect of diffracted and radiated waves from the structure. However, the GS semi-empirical distribution does not have this limitation since the parameters are updated empirically utilizing the disturbed wave records. In the following section, the performance of the GS semi-empirical model on

wave records measured in the vicinity and underneath the mini-TLP model test is evaluated.

IV.3.6.2 Disturbed waves

Similar to Table 15, the first four moments and L-moments of the measured waves at A2, A3, and A4 of the mini-TLP model test during the seastates S-1 and S-2 are presented in Table 17. The sample name in this table indicates the sea-state and air-gap probe at which the sample is measured, e.g. the wave record measured at A2 during the seastate S-1 is called S1A2. The statistics in this table are from a single 3-hr realization of wave records measured for the fixed structure configuration. Considering the relatively small structural oscillations, it is expected that the contribution of radiated waves being insignificant as compared to the contributions of incident and diffracted waves. The advantage of using the measurements of fixed structure configuration is that the wave data are automatically decoupled from the structural motions.

Comparing the statistics given in Table 15 for undisturbed waves in S-1 and the corresponding statistics of the disturbed waves in Table 17, it can be seen that the wave characteristics in the vicinity and underneath of the structure are not significantly changed. However, the wave elevations are significantly modified in the vicinity of the mini-TLP during the seastate S-2. In both seastates, the second moments (L-moments) of the wave measurements at A2, A3, and A4 show an increase as compared to the second moment of the undisturbed waves. This is mainly due to the contribution of diffracted waves to the wave energy spectrum. The contribution of diffracted waves decreases from air-gap probe A2 to A4, as expected. Comparing to the statistics of the corresponding

undisturbed waves, the third and fourth moments (L-moments) of the wave measurements during the seastate S-1 are not significantly modified while the situation is different for the measurements during the S-2. It is observed in both seastates that waves are more non-linear over the pontoons, i.e. A2 and A4, as compared to the waves at the middle of the moon-pool, i.e. A3. An important observation here is that the estimates of skewness and kurtosis of the wave records S2A2 (see Table 17), 0.422 and 0.808, respectively, are significantly higher than the corresponding statistics of the incident waves S2W (see Table 15), 0.288 and 0.099. The similar pattern is not observed in the L-skewness and L-kurtosis of this measurement which will be discussed later in this section.

Table 17 L-moments and moments of normalized disturbed surface elevation.

Estimate	Disturbed wave elevation samples					
	S1A2	S1A3	S1A4	S2A2	S2A3	S2A4
$l_1(\zeta_n)$	0.000	0.000	0.000	0.000	0.000	0.000
$l_2(\zeta_n)$	0.630	0.627	0.608	0.651	0.610	0.598
$\hat{\tau}_3(\zeta_n)$	0.004	0.009	0.004	0.054	0.035	0.056
$\hat{\tau}_4(\zeta_n)$	0.133	0.128	0.125	0.127	0.117	0.121
$\hat{\mu}_1(\zeta_n)$	0.000	0.000	0.000	0.000	0.000	0.000
$\hat{\mu}_2(\zeta_n)$	1.258	1.243	1.167	1.355	1.168	1.128
$\hat{s}_{\zeta_n}$	0.029	0.062	0.026	0.422	0.202	0.328
$\hat{K}_{\zeta_n}$	0.200	0.144	0.154	0.808	-0.048	0.088

Utilizing the moments and L-moments given in Table 17 the numerical values of the GS semi-empirical model parameters estimated utilizing MoM and MoL for the disturbed wave samples are given in Table 18. Once more, the estimates of the two parameter estimation methods are fairly close. The only exceptions are the estimates of β and γ for the wave record S2A2 where the MoM has captured a significantly larger second-order contribution and absolute shifting term. The considerably larger parameter α than 1.0 of the measurements at A2 and A3 denotes the significant contribution of the diffracted waves in these wave records. The second-order effects are found to be considerably larger in the wave records measured during the second design seastate S-2 especially in the measurements over the pontoons i.e. S2A2 and S2A4.

Table 18 Empirically estimated parameters of Gaussian-Stokes model (disturbed wave samples).

Estimate	Method	Disturbed wave elevation samples					
		S1A2	S1A3	S1A4	S2A2	S2A3	S2A4
$\hat{\alpha}$	MoL	1.117	1.112	1.077	1.153	1.081	1.060
	MoM	1.122	1.115	1.080	1.158	1.079	1.059
$\hat{\beta}$	MoL	0.005	0.010	0.004	0.064	0.039	0.061
	MoM	0.005	0.012	0.005	0.082	0.036	0.058
$\hat{\gamma}$	MoL	-0.005	-0.010	-0.004	-0.064	-0.039	-0.061
	MoM	-0.005	-0.012	-0.005	-0.082	-0.036	-0.058

The exceedance probability distributions of the wave sample S1A2 are presented in Figure 26. In this figure the distributions of the GS model, as well as, the empirical Gaussian and beta are given. As shown in this figure, the wave record S1A2 can be reasonably approximated as a linear process and the semi-empirical models as well as empirical Gaussian distributions are successful in capturing the probability distribution of this data set. The RMSE distributions of the quantile estimates of the semi-empirical and empirical distributions for the sample S1A2 are given in Figure 26 (b). The error of the semi-empirical model remains less than 0.2 (equivalent to 0.2 (m)) for the entire distribution. The situation is significantly different for the wave record S2A2; see Figure 27. In this case the sample distribution shows a sudden slope change at probability of $P = 2.5\times10^{-3}$ that cannot be explained by second-order wave theory. The physical phenomenon is similar to the hydraulic jump in the open channel flow where a sudden elevation change happens at a critical flow velocity. It can be argued that the observations with $P < 2.5\times10^{-3}$ do not follow the same probability distribution as the rest of the sample and can be removed from the sample. The distribution change is not captured by the GS models which results in significant quantile underestimation. The sudden change in the sample tail distribution is the reason of significantly high third and fourth sample moments of the wave sample S2A2 (see Table 17). The L-moments are less influenced by these unexpectedly large values and consequently do not show the same pattern. The RMSE distributions of the quantile estimates for this example are given in Figure 27 (b). It can be seen that for this data set, the GS model is only applicable for quantiles with $P > 2.5\times10^{-3}$.

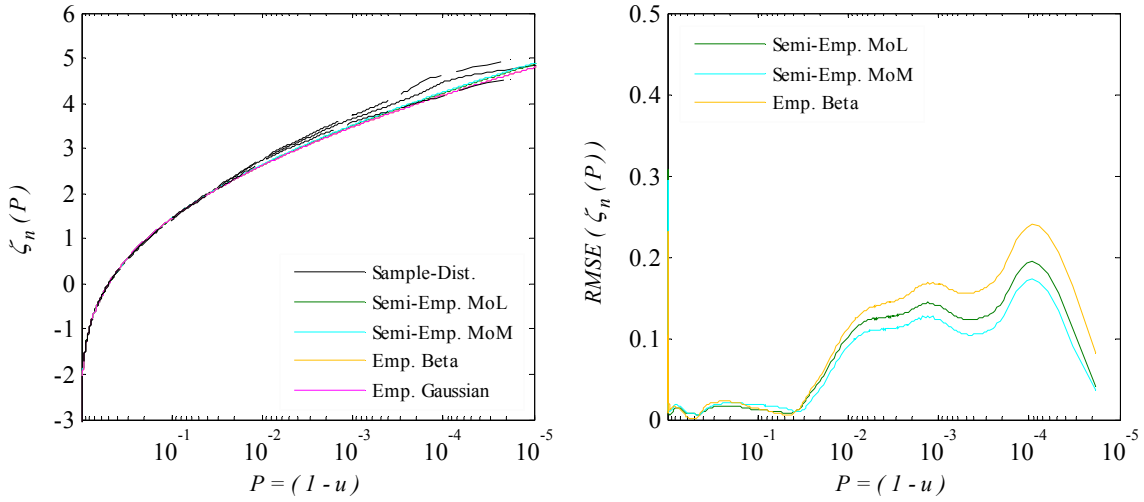


Figure 26 (a) Exceedance probability distributions of the disturbed wave sample S1A2 (b) RMSE distributions of the quantile estimates.

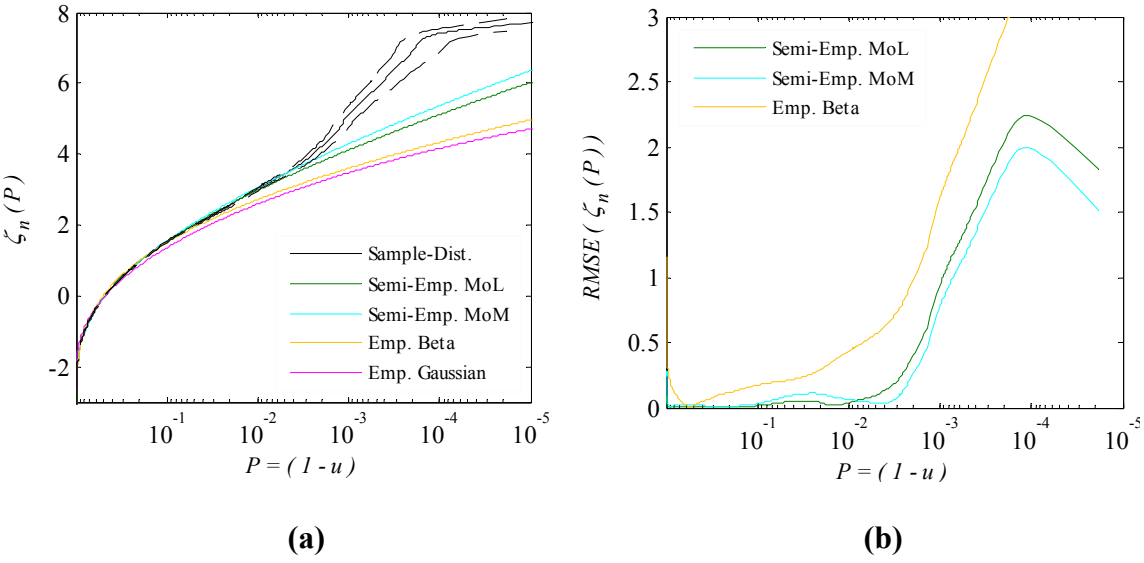


Figure 27 (a) Exceedance probability distributions of the disturbed wave sample S2A2, (b) the RMSE of the quantile estimates.

Similarly in Figure 28 and Figure 29, the exceedance probability distributions and RMSE distributions of the quantile estimates for the samples S1A3 and S2A3 are presented, respectively. The general conclusion here is that the GS semi-empirical model is robust in estimating the probability distribution of data and performs better than the empirical beta distribution for the most part of the distribution. However, it seems that the GS model tends to overestimate the tail distribution where the sample tail becomes flatter ($P \approx 3\times10^{-4}$). The situation is more critical for waves measured during the more energetic seastate S-2, where the GS model has captured a larger second-order contribution resulting in a heavier positive tail. The sample distribution slope change might be caused by an energy loss mechanism, e.g. wave breaking, or a balance between the positive and negative non-linear terms which neither was considered in the GS mathematical model, i.e. perturbation expansion. A similar pattern has been observed in the exceedance probability distribution of the wave elevation samples S1A4 and S2A4 shown respectively in Figure 30 and Figure 31.

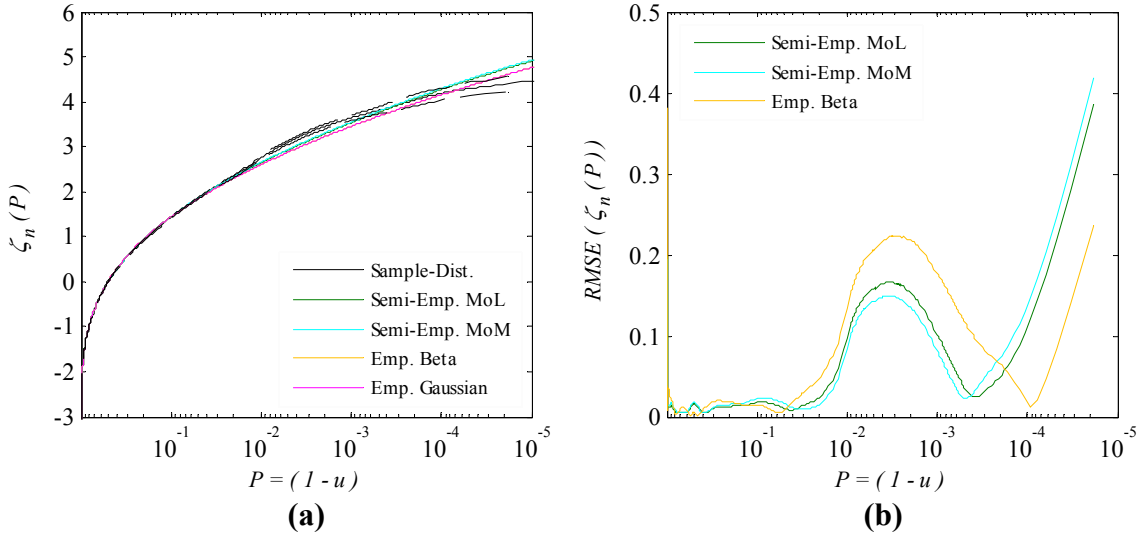


Figure 28 (a) Exceedance probability distributions of the disturbed wave sample S1A3, (b) RMSE distributions of the quantile estimates.

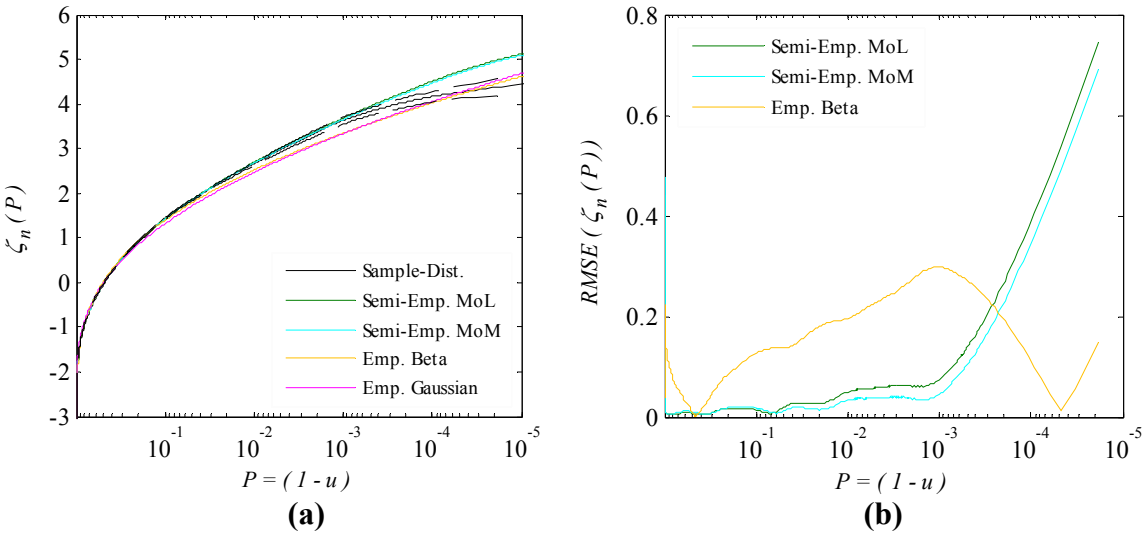


Figure 29 (a) Exceedance probability distributions of the disturbed wave sample S2A3, (b) RMSE distributions of the quantile estimates.

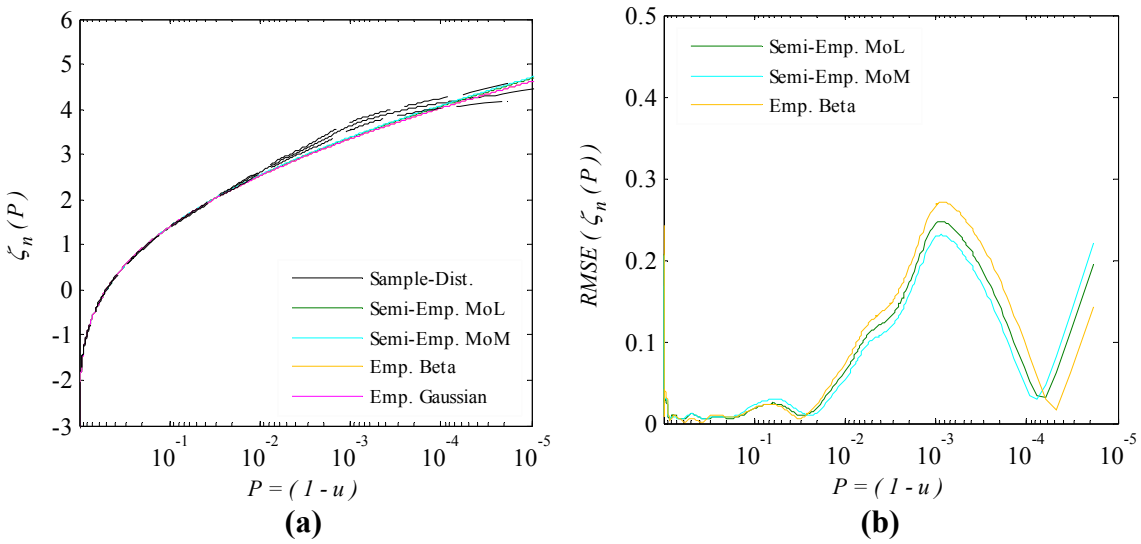


Figure 30 (a) Exceedance probability distributions of the disturbed wave sample S1A4, (b) RMSE distributions of the quantile estimates.

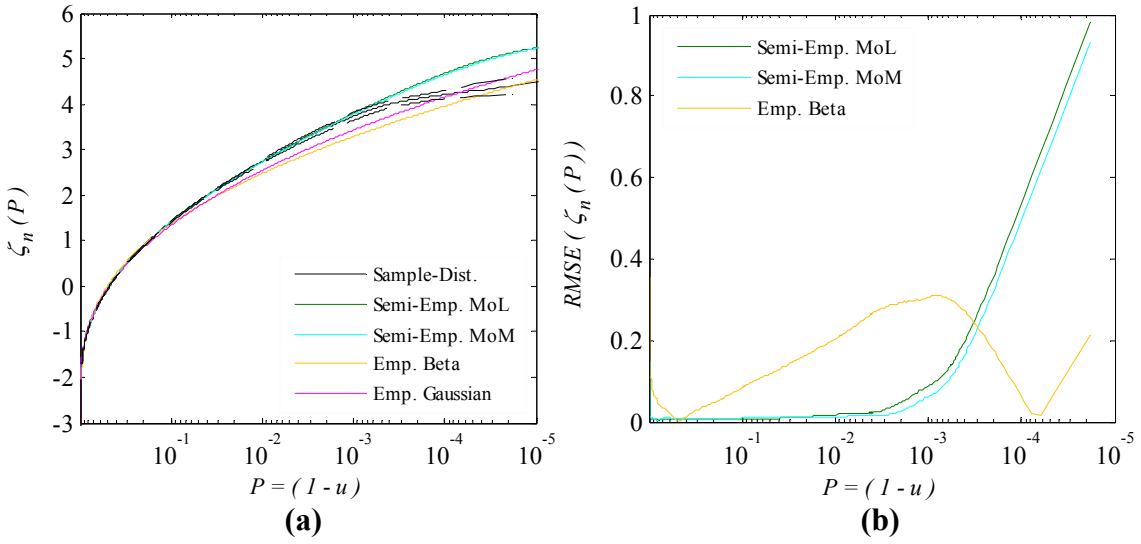


Figure 31 (a) Exceedance probability distributions of the disturbed wave sample S2A4, (b) RMSE distributions of the quantile estimates.

IV.4 Probability Distribution of Wave Crests

IV.4.1 Introduction

The probability distribution of wave crests in the vicinity and beneath deepwater platforms presents a significant design problem for an ocean engineer specifying the deck elevation. Design codes require adequate air gap elevation to avoid potential impact loads on the underside of the deck and inundation of the topsides. For offshore platforms, vertical columns, submerged pontoons, diagonal bracings and exposed riser arrays can greatly influence the nonlinear nature of the design seas through diffraction and radiation effects.

The theoretical probability distributions of wave crests in the open ocean have been extensively studied in the last five decades. Longuet-Higgins [1] showed that the wave crests of a linear wave train follow Rayleigh distribution and then later on Cartwright and Longuet-Higgins [6] simplified the Rayleigh distribution for waves with narrow-band energy spectra. Naess [5] derived a new probability distribution for linear wave heights and crests which included the effect of the wave elevation spectrum width. Linear models were evaluated and found to be inadequate for modeling the probability distribution of large crests in extreme environments. Assuming that the wave surface elevation could be modeled as a narrow-band process, Tayfun [8] developed a theoretical model that described the transformation of a narrow-band linear crest distribution to a narrow-band nonlinear crest distribution. This model was subsequently modified by other researchers [13-15, 66, 67]. Based on Gram-Charlier theory Tayfun [12] and Al-Humoud et al. [68] developed a nonlinear crest distribution function resulting in a model

that was less restrictive with regards to the spectral properties of ocean wave elevations. It was shown that the performance of the new model was better than the linear model but that the accuracy of the predictions remained somewhat elusive. Jha and Winterstein [57] introduced a second order transfer function of linear narrow-band wave crests to nonlinear wave crests based on Hermite model and showed that the model performs well for numerically generated second order waves. Extending Boccotti's quasi-determinism theory for second order waves, Fedele and Arena [16] derived a new probability distribution for nonlinear wave crests that is valid for deep water waves with finite bandwidth. It should be noted that the predictions of various nonlinear models yield nearly identical results for large waves in the deepwater limit.

As an alternative to theoretical models, Forristall [18] utilized a two-parameter Weibull distribution as the basis to estimate the probability distribution of wave heights empirically. He later related his Weibull model parameters to wave steepness and the Ursell parameter providing a different perspective for estimating the probability distribution of wave crests [19]. That improved model was more accurate for large wave crests but was shown to be less so for small waves [69]. Stansell [20, 21] analyzed surface elevation measurements collected during severe storms in the North Sea and utilized Generalized Pareto Distribution (GPD) to model the probability distribution of extreme wave crests. The results of his studies indicated that the GPD predicts the statistics of the extreme crests and troughs more accurately than the Rayleigh distribution.

Moment based Hermite model was used to model non-Gaussian behavior of waves interacting with floating structures [70-72]. In these studies the moments of surface wave elevation were estimated from different combinations of Stokes wave theory and linear and non-linear diffraction analysis. The results of the statistical models were compared with the measurements from a model test study of a semi-submersible platform. The comparisons indicated that the linear diffraction models were non-conservative and under-predict the air gap demand. Although the models based on second-order diffraction analysis were expected to better reflect observed data, it was found that these radiation/diffraction panel calculations sometimes over-predict air gap demand [70, 71]. As an alternative, Sweetman and Winterstein [72] applied a hybrid model which combines the first order diffraction results and the second order Stokes' waves. The new hybrid model was computationally simpler than the model based on the complete second order diffraction analysis and the accuracy of the model in predicting the air-gap was comparable. Sweetman and Winterstein [72] also applied the narrow-band nonlinear model derived earlier by Tayfun [8] and showed that the model results were reasonable considering the simplicity of the model. These studies have given significant insight to the extreme statistics of waves in vicinity and underneath of floating structures while the model predictions for extremely complex wave fields close to vertical structural columns and submerged pontoons still need improvement.

In this research study the three-parameter Rayleigh-Stokes (RS) and the four-parameter Weibull-Stokes (WS) semi-empirical models are utilized to estimate the probability distribution of ocean wave crests as well as wave crests in the vicinity and

beneath an offshore structure. The application of semi-empirical models for wave crests in the vicinity of offshore structures was first presented by Izadparast and Niedzwecki [73]. It was shown that the model is robust in capturing the probability distribution of mildly non-linear wave crests. Following a similar approach, here the RS and WS models' performance is evaluated over the mini-TLP model test measurements and numerically simulated waves. The semi-empirical model results are compared with appropriate theoretical and empirical models in case of undisturbed ocean wave crests. Additionally, the efficiency of the parameter estimation methods, i.e. method of L-moments (MoL) and method of moments (MoM) in capturing the probability distribution of sample data sets are evaluated.

Moreover, the statistics of extreme wave crests, i.e. wave crest maxima, in the area close to the offshore platform is studied. Knowing the probability distribution of wave crests, the probability distribution of crest maxima at a single point can be estimated utilizing extreme value theory (see section II.6) assuming that wave crests are independent random variables. The assumption of independent wave crests is not theoretically justified knowing that wave crests are commonly positively correlated. However, Krogstad and Barstow [74] utilized simulated samples to show that the wave crests correlation can be neglected in extreme analysis for large number of waves $N > 100$. In the open ocean, the Rayleigh distribution of linear wave crests has been widely used to approximate the probability distribution of wave crest maxima. However, it is well understood that the linear approximation underestimates the crests in extreme environments where nonlinear effects play an import role. Nerzic and Prevosto

[75] included the nonlinear effects in calculation of crest maxima utilizing a theoretical representation of the third-order RS model. The results of their study show a reasonable agreement between the model statistics and the statistics obtained from measured waves. Later on, the theoretical RS model was used by [74, 76, 77] to estimate the probability distribution of crest maxima. The Forristall's 2-parameter Weibull distribution is also used to estimate the probability distribution of crest maxima [74, 78]. Here, the probability distribution of disturbed wave crest maxima is estimated by application of RS and WS semi-empirical models and the statistics are compared with the sample estimates.

IV.4.2 Theoretical probability distributions

Longuet-Higgins [1] showed that the probability distribution of linear wave crests can be approximated by Rayleigh distribution defined in Eq. (2.42) with the Rayleigh parameter $R = (a_{rms}^2 / \sigma^2)/2$ where a_{rms} is the root-mean-square of the crest-height. For the case of linear narrow-banded waves $a_{rms} = \sqrt{2}\,\sigma_\eta$ and subsequently the Rayleigh parameter has a value of unity for normalized waves [6]. Later, Longuet-Higgins [79] showed that the nonlinear effects tend to increase the ratio of $\left(a_{rms}^2 / \sigma_\eta^{\,2}\right)/2$ while spectrum width effects may either increase or decrease the ratio. The effect of the combination of nonlinearity and spectrum width on the value of R is not clear from a theoretical perspective, but what is important is that the value of the Rayleigh parameter can be influenced to some extent by these effects. The Rayleigh distribution of linear

narrow-banded waves with $R = 1.0$ has been widely used to model the wave crests of intact ocean waves and is called the theoretical Rayleigh model in this document.

Utilizing Eq. (4.8) and (4.9) the wave crests of narrow-banded second-order modulated Stokes-waves can be represented as [8],

$$a_n = a + \frac{1}{2}\bar{k}a^2 \tag{4.28}$$

The non-dimensional form of the equation is obtained from normalizing the wave crests by the first-order standard deviation of waves $\sigma_\eta = m_0^{1/2}$, particularly,

$$\zeta_n = \zeta + \frac{1}{2}\bar{k}\sigma_\eta\zeta^2 \tag{4.29}$$

where, $\zeta_n = a_n/\sigma_\eta$ and $\zeta = a/\sigma_\eta$. The quadratic transformation in (4.29) is similar to the perturbation expansion used in development of the RS and WS semi-empirical models Eq. (2.3) with parameters $\gamma = 0.0$, $\alpha = 1.0$, and $\beta = \frac{1}{2}\bar{k}\sigma_\eta$. Assuming that linear wave crests follow the Rayleigh distribution with $R = 1.0$, Tayfun [8] derived the theoretical probability distribution of narrow-banded non-linear wave crests in deepwater which has a similar structural form as of the RS distribution given in Eq. (2.43). More recently, Tayfun [13] modified the model parameter for more general water depth by relating the parameter β to the Foristall's [19] Weibull model parameters. The statistics of these models, however, are almost identical for deepwater condition. Although the model was originally developed for wave crests of narrow-banded waves, it can be used for large crests of wave processes with finite spectrum width [13, 16].

IV.4.3 Empirical probability distributions

Rayleigh distribution, Eq. (2.42), is perhaps the most famous probability distribution in the field of ocean engineering. The only parameter of this distribution can be estimated from its relation with the first sample moment and L-moment as,

$$\hat{R} = \frac{2\hat{\mu}_1(\zeta_n)}{\pi} = \frac{2l_1(\zeta_n)}{\pi} \tag{4.30}$$

A more robust estimate of the Rayleigh parameter, i.e. maximum likelihood estimate, is obtained from the definition of the sample root-mean-square, specifically,

$$\hat{R} = \frac{1}{2N}\sum_{i=1}^{N} {\zeta_{n,i}}^2 \tag{4.31}$$

where N is the sample size. The empirical Rayleigh distribution could capture the effects of non-linearity and spectrum width on the probability distribution of the wave crests to some extent [79]; however, it has limited flexibility in capturing the probability distribution of data. It should be noted that Rayleigh distribution has constant skewness and excess kurtosis of 0.631 and 0.245, respectively, that are independent of the parameter R. Similarly, the L-skewness and L-kurtosis of Rayleigh distribution, 0.114 and 0.105 are independent of the model parameter.

The two-parameter Weibull distribution is a more general formulation of the Rayleigh distribution and has been widely used to model the probability distribution of wave crests. Forristall [18, 19] utilized the distribution to model the probability distribution of ocean wave crests and Niedzwecki et al. [22] used the empirical Weibull distribution to estimate the probability distribution of measured wave crests in the vicinity of offshore structures. The PDF of the Weibull distribution with shape κ and

scale ϑ parameters is given in Eq. (2.55). The estimates of the Weibull parameters can be obtained from MoL and MoM, respectively,

$$\hat{\kappa} = \frac{\ln(2)}{\ln\left(\frac{l_1(\zeta_n)}{l_1(\zeta_n) - l_2(\zeta_n)}\right)}, \qquad \hat{\vartheta} = \frac{l_1(\zeta_n)}{\Gamma\left(1 + 1/\hat{\kappa}\right)}. \tag{4.32}$$

and,

$$\frac{\hat{\mu}_2(\zeta_n)}{\hat{\mu}_1{}^2(\zeta_n)} = \frac{\Gamma\left(1 + \frac{2}{\hat{\kappa}}\right)}{\Gamma^2\left(1 + \frac{1}{\hat{\kappa}}\right)} - 1, \qquad \hat{\vartheta} = \frac{\hat{\mu}_1(\zeta_n)}{\Gamma\left(1 + 1/\hat{\kappa}\right)}. \tag{4.33}$$

An iterative scheme is required to solve the first relation in Eq. (4.33) for $\hat{\kappa}$.

IV.4.4 Sample data sets

The timeseries utilized here are the ones used in the wave elevation analysis described in IV.3.5. The wave crests are defined as the maximum wave elevation in between each two consecutive zero up-crossings. For each wave record, the wave crest sample is developed from zero-crossing analysis of the three-hour (prototype scale) wave elevation timeseries. The probability distribution of undisturbed wave crests measured at the reference probe RF6 (see Figure 19), and disturbed wave crests measured at air gap probes A2-A4 are studied here. Additionally, the wave crests distribution of the numerically generated second-order wave record (S-3) is studied.

IV.4.5 Analysis and results

IV.4.5.1 Undisturbed wave crests

The sample estimates of the first four L-moments and moments of the normalized undisturbed wave crests $\zeta_n = a/m_0^{1/2}$ in the three seastates S-1, S-2, and S-3 (see Table 14) are presented in Table 19. For simplicity the crest sample measured at RF6 during the seastate S-1, the crest sample measured at RF6 during the seastate S-2, and the crest sample of numerically generated second-order waves from the seastate S-3 are called S1C, S2C, and S3C, respectively. Additionally in Table 19, the L-moments and moments of the Rayleigh distribution of linear process as well as the statistics of Tayfun's theoretical RS model are given. It is observed that the linear model underestimates the first two moments (L-moments) of the crest samples especially the crest samples S2C and S3C. The Tayfun model's first two moments (L-moments) are better approximations of the sample estimates, while the second moments of wave crests in S2C and S3C samples are still underestimated. Both linear and Tayfun models overestimate the sample skewness which is more sensible in case of Tayfun model. The sample crest S1C and S3C have smaller kurtosis (L-kurtosis) than the Rayleigh distribution while the sample crests S2C has a larger kurtosis than the linear model. The Tayfun model consistently overestimates the sample kurtosis in the studied samples.

Table 19 L-moments and moments of normalized undisturbed wave crests.

Estimate	Linear Process	Undisturbed wave crest samples					
		S1C		S2C		S3C	
		Sample	Tayfun	Sample	Tayfun	Sample	Tayfun
$l_1(\zeta_n)$	1.253	1.273	1.277	1.465	1.357	1.432	1.465
$l_2(\zeta_n)$	0.367	0.457	0.379	0.482	0.419	0.506	0.473
$\hat{\tau}_3(\zeta_n)$	0.114	0.099	0.121	0.109	0.141	0.109	0.163
$\hat{\tau}_4(\zeta_n)$	0.105	0.094	0.107	0.119	0.113	0.102	0.119
$\hat{\mu}_1(\zeta_n)$	1.253	1.273	1.277	1.465	1.357	1.432	1.465
$\hat{\mu}_2(\zeta_n)$	0.429	0.658	0.460	0.741	0.570	0.812	0.740
$\hat{s}_{\zeta_n}$	0.631	0.564	0.673	0.627	0.799	0.605	0.937
$\hat{K}_{\zeta_n}$	0.245	0.034	0.349	0.281	0.685	0.145	1.104

Utilizing the sample statistics given in Table 19, the parameters of the RS and WS semi-empirical models are estimated with MoL and MoM, and the results for the studied samples are presented in Table 20. In estimation of the RS and WS parameters, without loss of generality, it is assumed that $R = 1.0$ and $\upsilon = \sqrt{2}$ to make the value of model parameters directly comparable with the theoretical values. The theoretical estimates of the RS model parameters α, β and γ obtained from Tayfun approximation are given in Table 21. For all the studied examples, the parameter estimates of MoL and MoM of both RS and WS models were found to be reasonably close.

Table 20 Empirically estimated parameters of Rayleigh-Stokes and Weibull-Stokes models (undisturbed normalized wave crest samples).

Parameter	Model	Method	Undisturbed wave crest samples		
			S1C	S2C	S3C
$\hat{\alpha}$	RS	MoL	1.327	1.343	1.412
	RS	MoM	1.306	1.318	1.404
	WS	MoL	1.331	1.348	1.421
	WS	MoM	1.307	1.295	1.409
$\hat{\beta}$	RS	MoL	-0.031	-0.011	-0.012
	RS	MoM	-0.023	-0.002	-0.010
	WS	MoL	-0.059	0.031	-0.028
	WS	MoM	-0.047	0.040	-0.032
$\hat{\kappa}$	RS	MoL	NA	NA	NA
	RS	MoM	NA	NA	NA
	WS	MoL	1.831	2.241	1.942
	WS	MoM	1.877	2.159	1.904
$\hat{\gamma}$	RS	MoL	-0.329	-0.195	-0.313
	RS	MoM	-0.317	-0.184	-0.308
	WS	MoL	-0.264	-0.258	-0.293
	WS	MoM	-0.272	-0.234	-0.271

Table 21 Theoretical estimates of the Rayleigh-Stokes model parameters (undisturbed normalized wave crest samples).

Parameter	Model	Undisturbed wave crest samples		
		S-1	S-2	S-3
$\hat{\alpha}$	Tayfun	1.000	1.000	1.000
$\hat{\beta}$	Tayfun	0.012	0.052	0.106
$\hat{\kappa}$	Tayfun	0.000	0.000	0.000

It has been observed that the empirical estimates of the parameter α are significantly larger than the theoretical value of unity. The large empirical value of α is caused by relatively large sample second moment (L-moment) and relatively small sample skewness (L-skewness). The large value of α has caused negative values of β in the semi-empirical models which are not consistent with the theoretical estimates of parameter β presented in Table 21. The negative second-order contribution balances the contribution of the first-order term especially on the tail of the distribution. The RS and WS models have captured negative shifting parameter γ indicating that the distributions are shifted towards the smaller values. In the theoretical model this parameter is assumed to be zero. The negative shifting parameter is caused by small amplitude waves oscillating at high frequencies which are expected in wave processes with finite spectrum width and are neglected in the narrow-banded models. Although the shifting parameter does not have a clear theoretical interpretation, it improves the performance of the semi-empirical models in capturing the probability distribution of data, especially the distribution of small crests. It should be noted that negative γ causes the RS and WS PDF to have non zero value for $\zeta_n \le 0$ which is not justified. To solve this problem, here the probability distributions are only estimated for $\zeta_n \ge 0$ assuming that $\text{Prob}(\zeta_n = 0) = \text{Prob}(\zeta_n \le 0)$.

The exceedance probability distributions of wave crest sample S1C are shown in Figure 32. In this figure the sample distribution is obtained from the semi-parametric method and is applied in the bootstrap analysis. The 95% confidence intervals of the

sample distribution are estimated utilizing 10,000 bootstrap samples. In Figure 32, the estimates of the theoretical models, i.e. Rayleigh distribution of linear process and Tayfun second-order approximation (Figure 32 (a)), semi-empirical models, i.e. RS and WS with parameters estimated by MoL and MoM (Figure 32 (b)), and empirical models, i.e. empirical Rayleigh and Weibull distributions (Figure 32 (c)), are compared. As shown in this figure, the linear model significantly underestimates the sample distribution. Although the Tayfun model improves the estimates of linear model, the model underestimates the large crests with small probability of exceedances. The semi-empirical models are robust in estimating the probability distribution of data and the estimates of different semi-empirical models found to be considerably close. The statistics of empirical Rayleigh and Weibull distributions are fairly accurate for crests with $P > 0.1$ while their accuracy is questionable for extreme crests. As shown in Figure 32 (c) empirical Rayleigh distribution underestimated the tail distribution, while the two-parameter Weibull distribution overestimated the extreme observations.

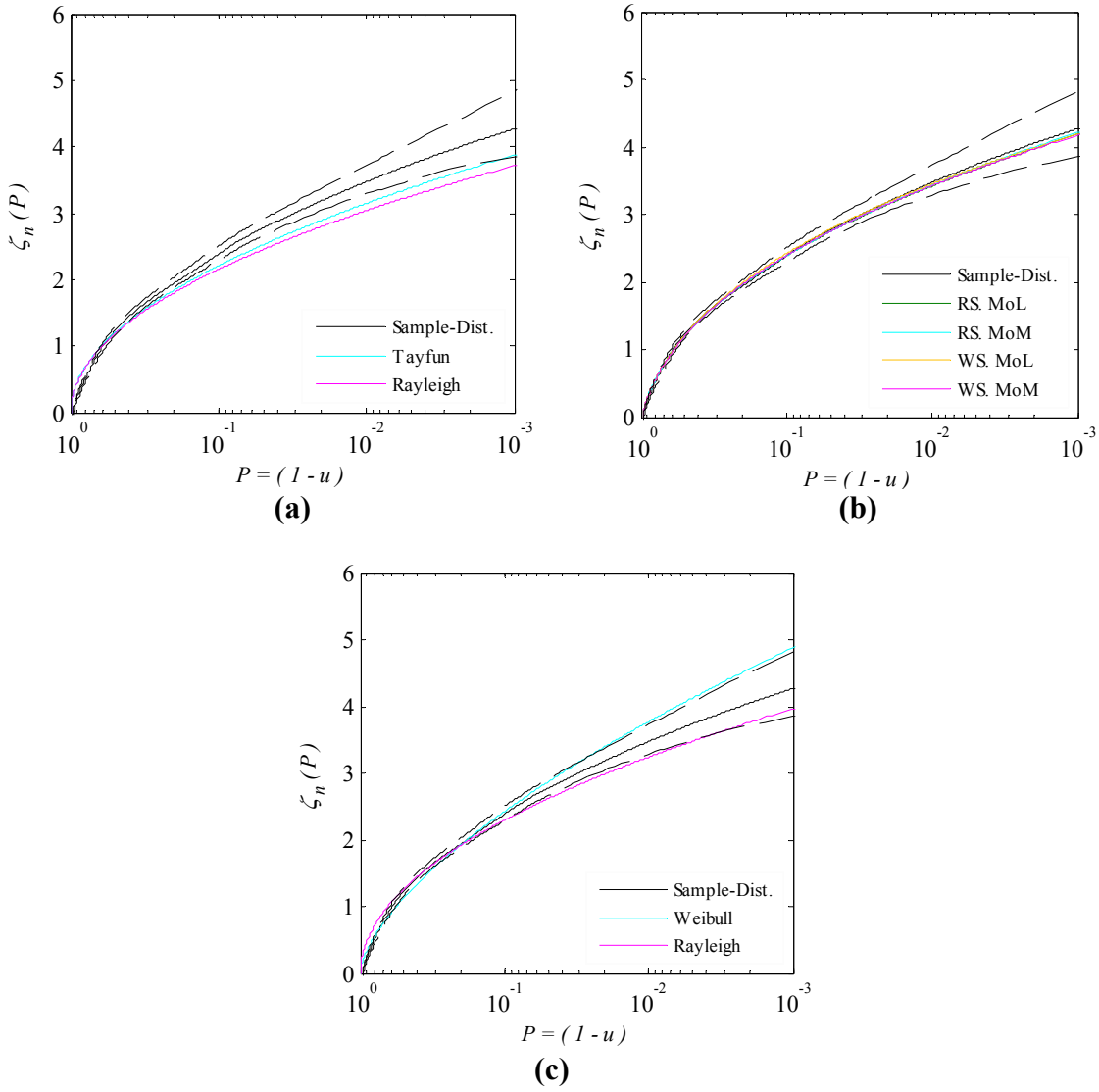


Figure 32 Exceedance probability distributions of wave crest sample S1C (a) theoretical models, (b) semi-empirical models, (c) empirical models. Dashed lines represent the 95% confidence limits.

Similar to Figure 32, in Figure 33 the exceedance probability distributions of wave crest sample S2C are presented. As shown in this figure, the theoretical models are not successful in representing the probability distribution of data. As expected, the Tayfun model provides a better estimate of the probability distribution while still underestimating the extreme observations. Similar to the previous case, the semi-empirical models were found to be considerably accurate in modeling the probability distribution of data. The empirical models perform better in this example as compared to their performance in the previous example shown in Figure 32 (c). However, similar to the previous example, the 2-parameter Weibull distribution tends to overestimate the tail distribution while the empirical Rayleigh distribution underestimates the extreme observations.

In the last example the probability distribution of numerically generated second-order wave crests S3C is studied. The exceedance probability distributions of normalized wave crests of this sample are shown in Figure 34. As expected, the Tayfun model performs reasonably well in this example. However, as shown in Table 19, the model overestimates the sample skewness and kurtosis and consequently has a heavier tail as compared to the sample distribution. As seen in previous examples for measured waves, the semi-empirical models are successful in capturing the probability distribution of simulated data and the estimates of MoL and MoM are reasonably close. In case of the empirical model the trend is similar to what was observed in the first example (Figure 32 (c)). It is observed that the empirical Rayleigh and Weibull distributions perform well for crests with $P > 0.1$ while both are not able to model the tail distribution accurately.

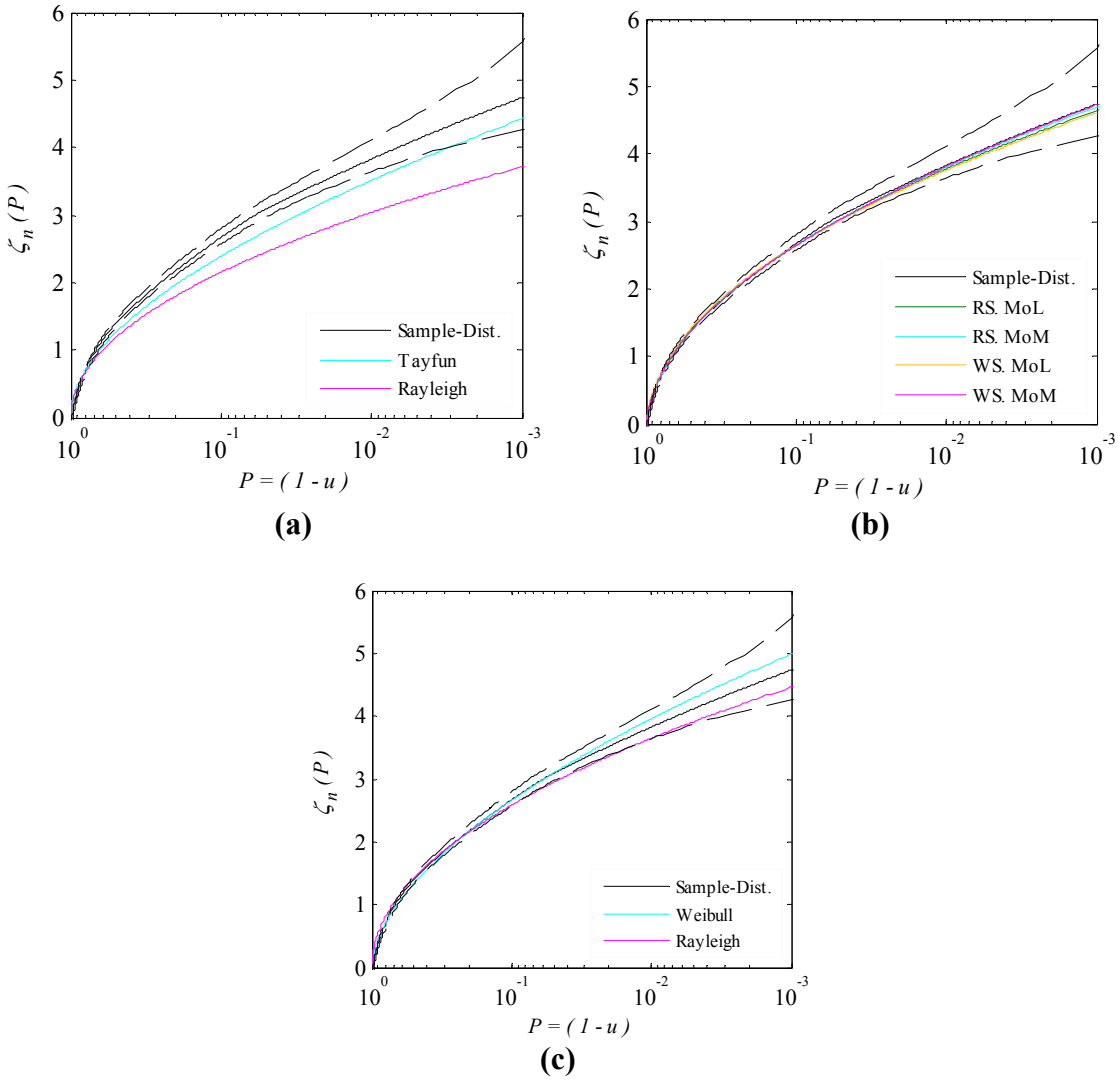


Figure 33 Exceedance probability distributions of wave crest sample S2C (a) theoretical models, (b) semi-empirical models, (c) empirical models. Dashed lines represent the 95% confidence limits.

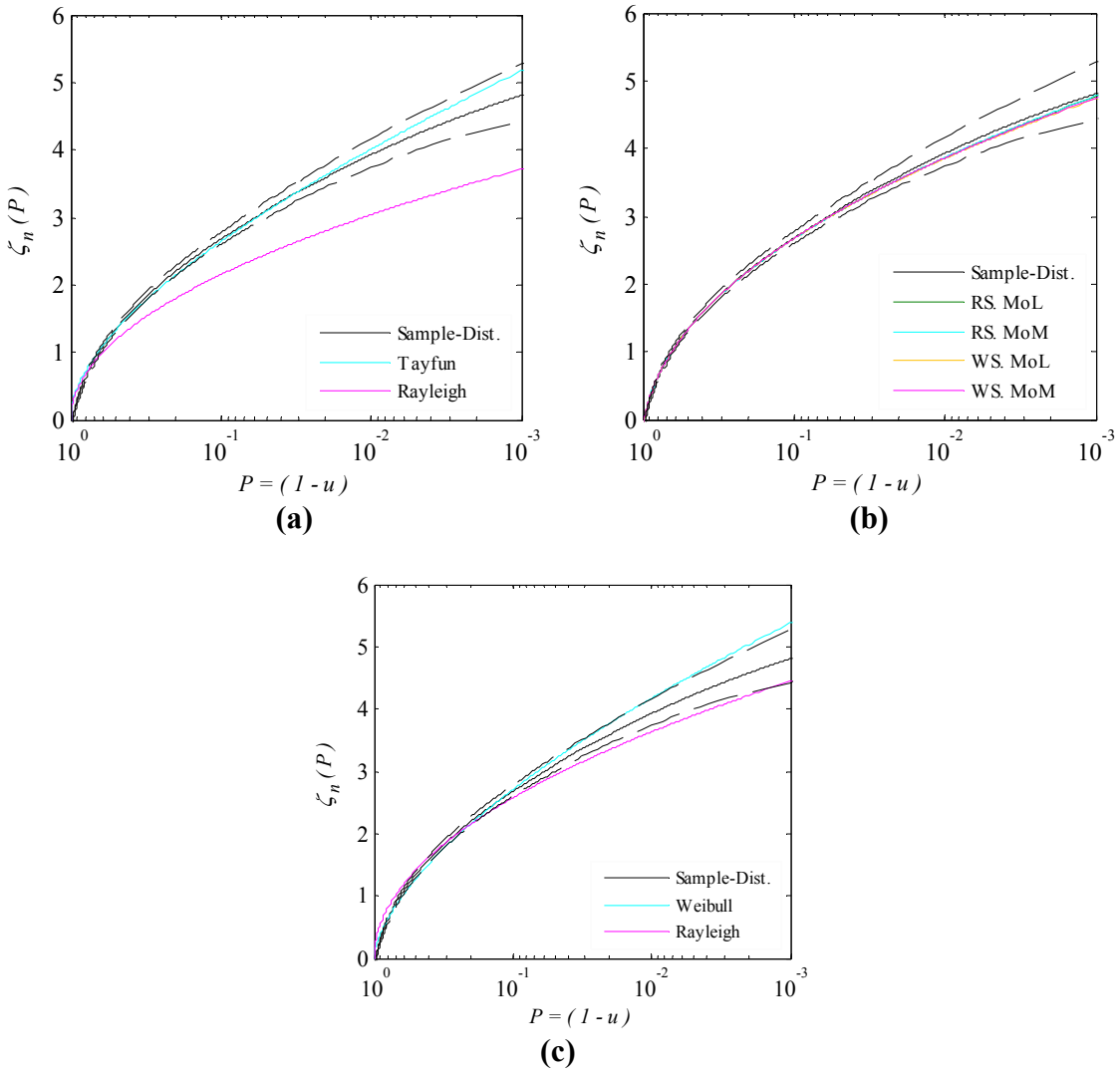


Figure 34 Exceedance probability distributions of wave crest sample S3C (a) theoretical models, (b) semi-empirical models, (c) empirical models. Dashed lines represent the 95% confidence limits.

Regarding these examples, one can clearly see the robustness of the semi-empirical models in capturing the probability distribution of undisturbed measured and simulated wave crests. The semi-empirical models constantly perform better as compared to the routinely used theoretical and empirical models. As shown in these examples the estimates of the RS and WS semi-empirical models with the model parameters estimated by MoL and MoM are reasonably close. In order to evaluate the performance of the semi-empirical models quantitatively, the RMSE of the quantile estimates are evaluated utilizing the bootstrap samples. The results of quantile error analyses for the three examples of undisturbed wave crests, i.e. crest samples S1C, S2C, and S3C, are presented in Figure 35. As shown in this figure the error in the quantile estimates of the semi-empirical models remains considerably small for the major part of the distribution with $P > 0.1$. As expected, the error in the model estimates increases in the tail of the distribution; however, in the three examples studied here, the error of the quantile estimates remains in an acceptable range. An interesting observation here is that the errors in the estimates of the four-parameter WS and three-parameter RS models with the parameters estimated from the same method follow a similar pattern. Additionally, it is observed that the distributions estimated by MoM perform slightly better in representing the probability distribution of extreme observations as compared to the same distributions estimated by MoL.

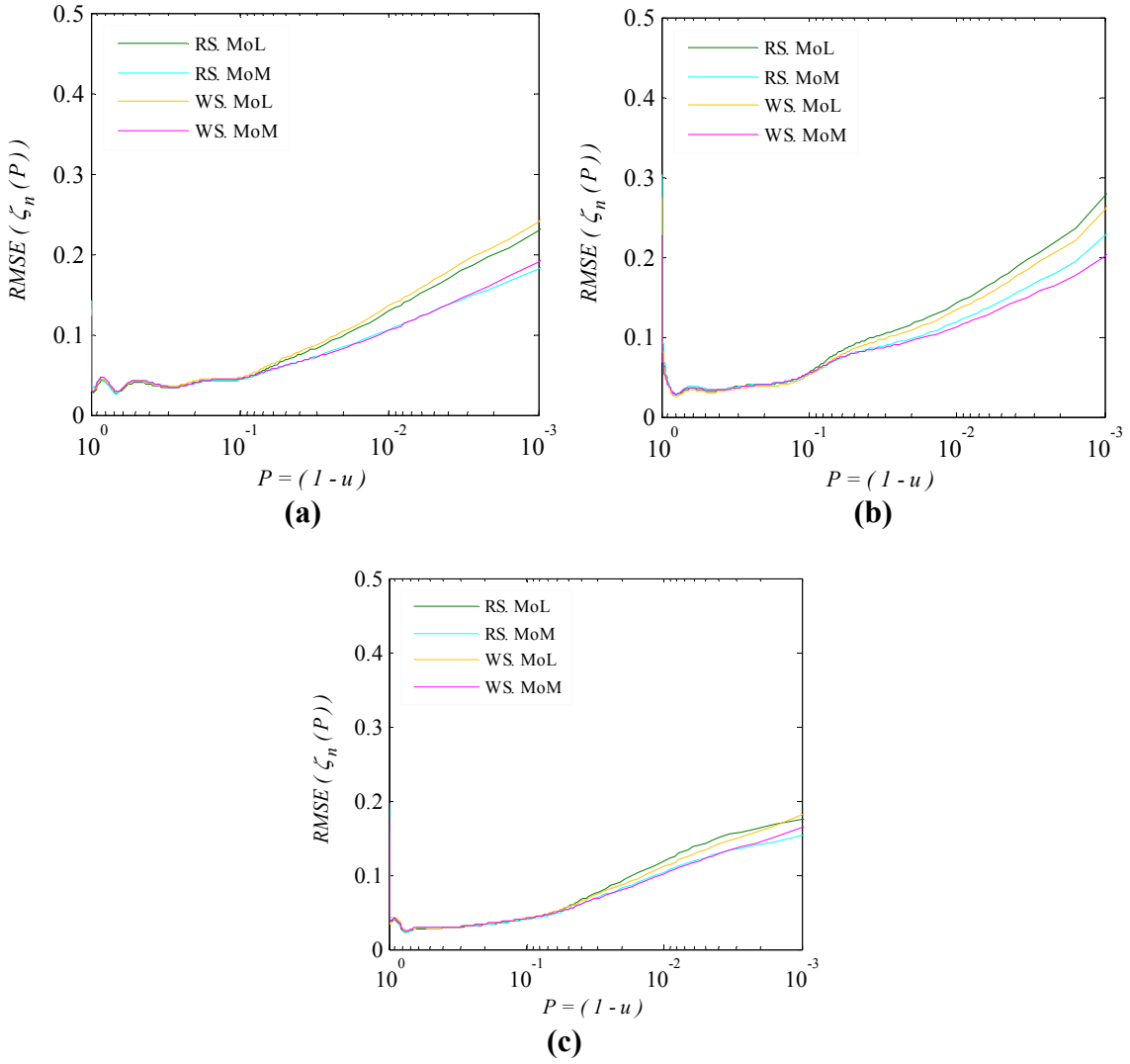


Figure 35 RMSE distribution of the semi-empirical models quantile estimates, (a) crest sample S1C, (b) crest sample S2C, (c) crest sample S3C.

In the final part of this section, the performance of the semi-empirical models in estimating the probability distribution of wave crest maxima is evaluated. For this purpose the probability distribution of wave crest maxima is estimated following the extreme value theory described in section II.6. In Figure 36 the PDFs of normalized wave crest maxima, $\zeta_{\max} = a_{\max}/\sigma_{\eta}$, in $N = 1,000$ waves estimated by application of the semi-empirical models are compared with the sample crest maxima distributions of the three samples S1C, S2C and S3C. As shown in Figure 36, the semi-empirical models are reasonably accurate in representing the sample extreme statistics. It should be noted that the sample distribution used here as the reference distribution contains a level of uncertainty which justifies the difference between the semi-empirical model PDFs and the sample distribution. The estimates of expected crest maxima in $N = 1,000$ waves for the three undisturbed wave crest samples are compared in Table 22. As shown in this table, the estimates of expected crest maxima are sensitive to the waves' spectrum characteristics. It is observed that the semi-empirical models are reasonably accurate in estimating the extreme statistics. This once more indicates that semi-empirical models are successful in capturing the overall probability distribution of undisturbed wave crest data sets as well as representing the tail distribution.

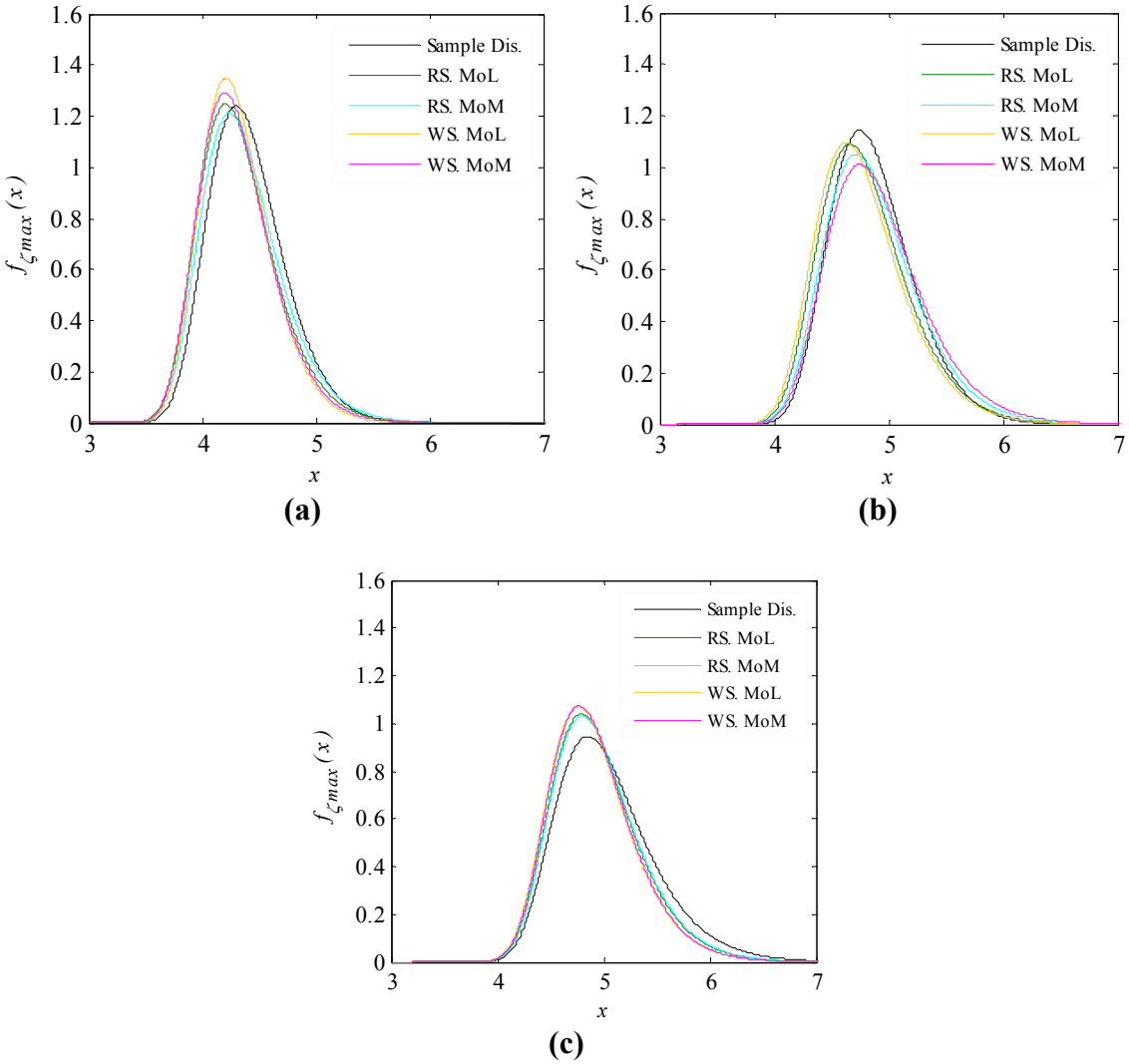


Figure 36 Probability density functions of undisturbed normalized wave crest maxima in 1,000 waves, (a) wave crest sample S1C, (b) wave crest sample S2C, (c) wave crest sample S3C.

Table 22 Estimates of normalized expected crest maxima in 1,000 waves of undisturbed wave samples.

Model	Undisturbed wave crest samples		
	S1C	S2C	S3C
Sample Dist.	4.40	4.91	5.04
RS. MoL	4.33	4.81	4.95
RS. MoM	4.37	4.88	4.96
WS. MoL	4.31	4.95	4.91
WS. MoM	4.31	4.92	4.92

IV.4.5.2 Disturbed wave crests

In this section the performance of the semi-empirical models, i.e. RS and WS models, in capturing the probability distribution of wave crests in the area close to an offshore structure is evaluated. For this purpose, the measured wave records at the air-gap probes during two seastates S-1 and S-2 of the mini-TLP model test are used. The data utilized here are from the rigid structure configuration and each data set models a three-hour seastate in prototype scale. In case of the benign sea condition, S-1, the measurements at the three air-gap probes A2, A3, A4 are analyzed while in case of the more energetic seastate S-2, only the measurements of the A3 and A4 probes are studied. The reason for excluding the wave record measured at A2 during S-2 is that the wave elevations and particularly the wave crests of this measurement do not follow the weakly nonlinear wave pattern and consequently the semi-empirical models are not applicable for this case.

The sample estimates of the four L-moments and moments of the disturbed wave crest samples are given in Table 23. For simplicity the samples are called with a

combination of the seastate and the air-gap probe at which the data are measured, e.g. wave crest sample measured during the seastate S-1 at air-gap probe A2 is called S1CA2. Each of the five samples studied here have their own distinctive statistical characteristics. Utilizing the sample statistics given in Table 23, the semi-empirical model parameters are estimated and the results are presented in Table 24. As shown in this table a reasonable agreement between the parameter estimates of different models in all the five examples is observed. The semi-empirical models have captured relatively large linear term contribution α while the contribution of the linear term is balanced with the negative second-order term contribution β. The negative β in these examples is caused by the relatively small third and fourth sample moments (L-moments) and considerably large second sample moments (L-moment) (see Table 23). Similar to what was observed in the undisturbed wave crest examples (see Table 20), the shifting parameter γ is consistently negative for the disturbed wave crest samples indicating that the sample distributions are shifted towards smaller wave crests. In all the studied examples, the estimates of WS's shape parameter κ are different from 2.0 which is the default value of the shape parameter in the RS model.

Table 23 L-moments and moments of normalized disturbed wave crests.

Estimate	Disturbed wave crest samples				
	S1CA2	S1CA3	S1CA4	S2CA3	S2CA4
$l_1(\zeta_n)$	1.326	1.349	1.305	1.484	1.369
$l_2(\zeta_n)$	0.491	0.478	0.461	0.486	0.516
$\hat{\tau}_3(\zeta_n)$	0.111	0.112	0.099	0.057	0.112
$\hat{\tau}_4(\zeta_n)$	0.083	0.101	0.100	0.099	0.075
$\hat{\mu}_1(\zeta_n)$	1.326	1.349	1.305	1.484	1.369
$\hat{\mu}_2(\zeta_n)$	0.760	0.726	0.676	0.736	0.836
$\hat{s}_{\zeta_n}$	0.595	0.631	0.601	0.367	0.563
$\hat{K}_{\zeta_n}$	0.022	0.151	0.224	-0.154	-0.162

Table 24 Empirically estimated parameters of Rayleigh-Stokes and Weibull-Stokes models (disturbed normalized wave crest samples).

Parameter	Model	Method	Disturbed wave crest samples				
			S1A2	S1A3	S1A4	S2A3	S2A4
$\hat{\alpha}$	RS	MoL	1.357	1.312	1.341	1.665	1.421
	RS	MoM	1.370	1.301	1.286	1.597	1.473
	WS	MoL	1.379	1.330	1.347	1.664	1.434
	WS	MoM	1.366	1.308	1.265	1.597	1.435
$\hat{\beta}$	RS	MoL	-0.007	-0.004	-0.031	-0.125	-0.005
	RS	MoM	-0.013	0.000	-0.011	-0.099	-0.026
	WS	MoL	-0.076	-0.037	-0.044	-0.063	-0.081
	WS	MoM	-0.066	-0.043	0.030	-0.022	-0.067
$\hat{\kappa}$	RS	MoL	NA	NA	NA	NA	NA
	RS	MoM	NA	NA	NA	NA	NA
	WS	MoL	1.691	1.867	1.944	2.313	1.682
	WS	MoM	1.730	1.795	2.168	2.391	1.788
$\hat{\gamma}$	RS	MoL	-0.359	-0.288	-0.314	-0.352	-0.401
	RS	MoM	-0.364	-0.282	-0.286	-0.319	-0.423
	WS	MoL	-0.207	-0.245	-0.295	-0.481	-0.194
	WS	MoM	-0.254	-0.206	-0.338	-0.476	-0.231

The exceedance probability distributions of the disturbed wave crest samples and the estimates of the semi-empirical and empirical models are compared in Figure 37 to Figure 41. As can be seen in these figures, the semi-empirical models are successful in representing the overall probability distribution of complex disturbed wave crest samples. However, the semi-empirical tail distributions are slightly different from the sample tails which is more sensible in samples measured during the second seastate, i.e. S2CA3 and S2CA4 (see Figure 40 and Figure 41). Once more, it should be mentioned that the sample tail distribution itself contains uncertainty and as shown in these figures the confidence limits become wider on the tails. The empirical Rayleigh and 2-parameter Weibull distribution, on the other hand, are not reliable in representing the probability distribution of data. The empirical Weibull distribution tends to significantly overestimate the large crests while the empirical Rayleigh distribution, except for sample S2CA3 (see Figure 40 (b)), underestimates the large disturbed crests.

The RMSE distribution of the semi-empirical models quantile estimates for the five disturbed crest samples are presented in Figure 42. The RMSE distributions shown in this figure are estimated utilizing the bootstrap analysis. As shown in this figure the error in the estimates of the semi-empirical models remains in an acceptable range even for extreme crests with small exceedance probabilities. In all the examples, the RS and WS models with parameters estimated by MoL and MoM perform similarly for the major part of the distribution with $P < 6\times10^{-2}$. The difference between the models happens on the tail part of the distribution where the sample tails are modified utilizing the GPD distribution. This indicates that a more accurate comparison between the

models tails requires longer data sets. Based on the distributions shown in Figure 42, it can be concluded that the MoM is performs slightly better in capturing the tail distribution of data as compared to the MoL. Additionally, in most examples the performance of the three-parameter RS model and four-parameter WS model with parameters estimated from a same parameter estimation method is comparable.

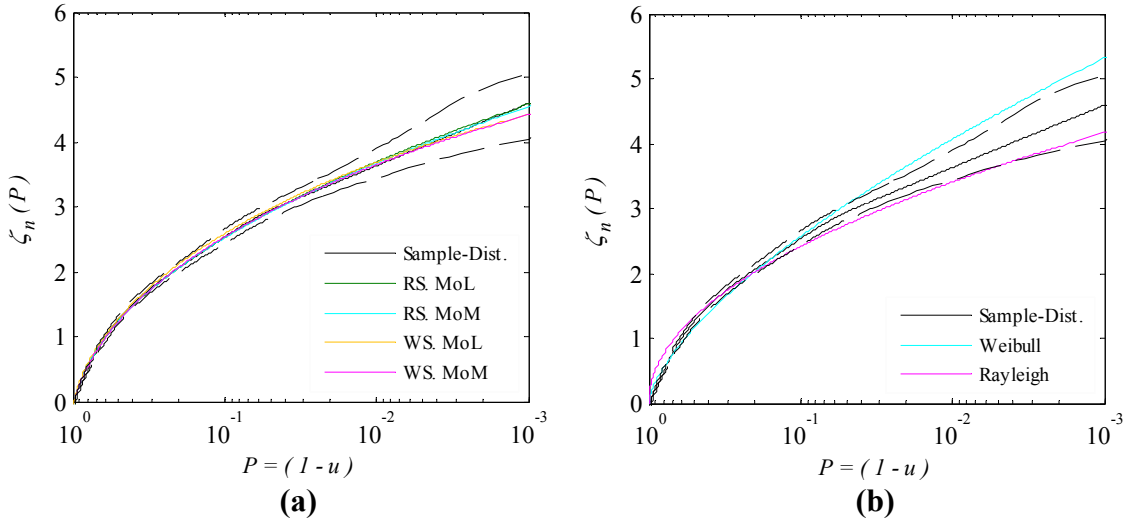


Figure 37 Exceedance probability distributions of normalized disturbed wave crest sample S1CA2 (a) semi-empirical models, (b) empirical models. Dashed lines represent the 95% confidence limits.

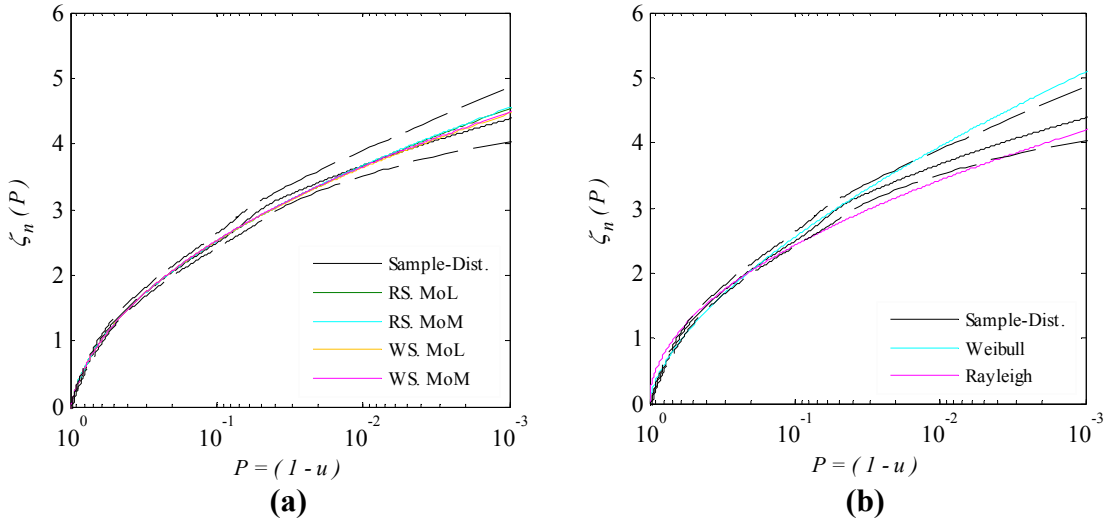


Figure 38 Exceedance probability distributions of normalized disturbed wave crest sample S1CA3 (a) semi-empirical models, (b) empirical models. Dashed lines represent the 95% confidence limits.

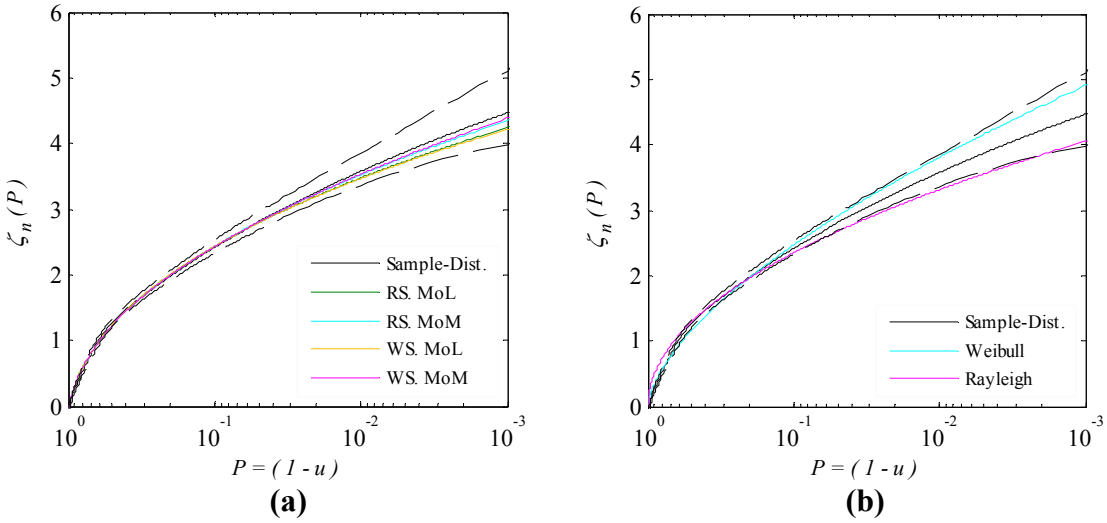


Figure 39 Exceedance probability distributions of normalized disturbed wave crest sample S1CA4 (a) semi-empirical models, (b) empirical models. Dashed lines represent the 95% confidence limits.

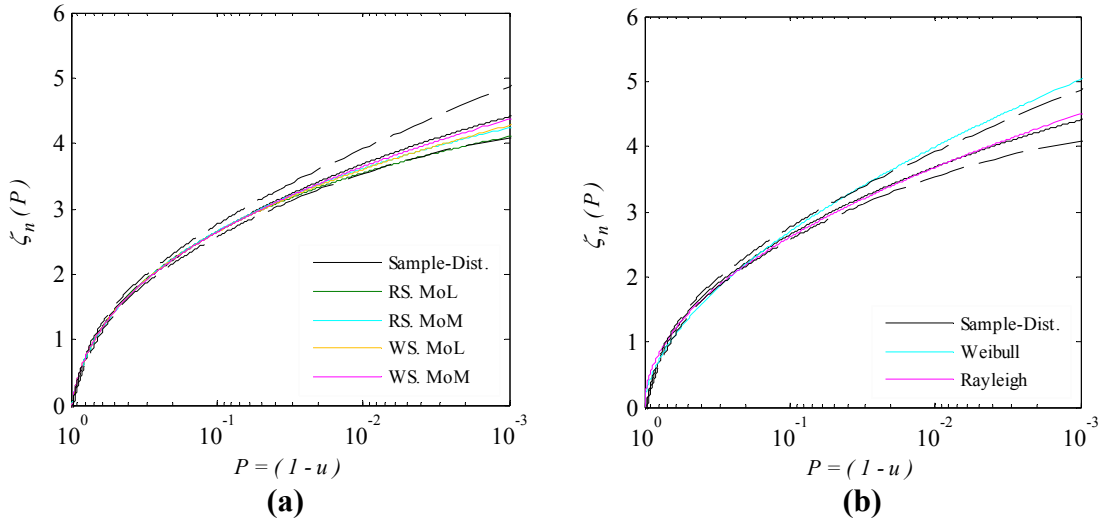


Figure 40 Exceedance probability distributions of normalized disturbed wave crest sample S2CA3 (a) semi-empirical models, (b) empirical models. Dashed lines represent the 95% confidence limits.

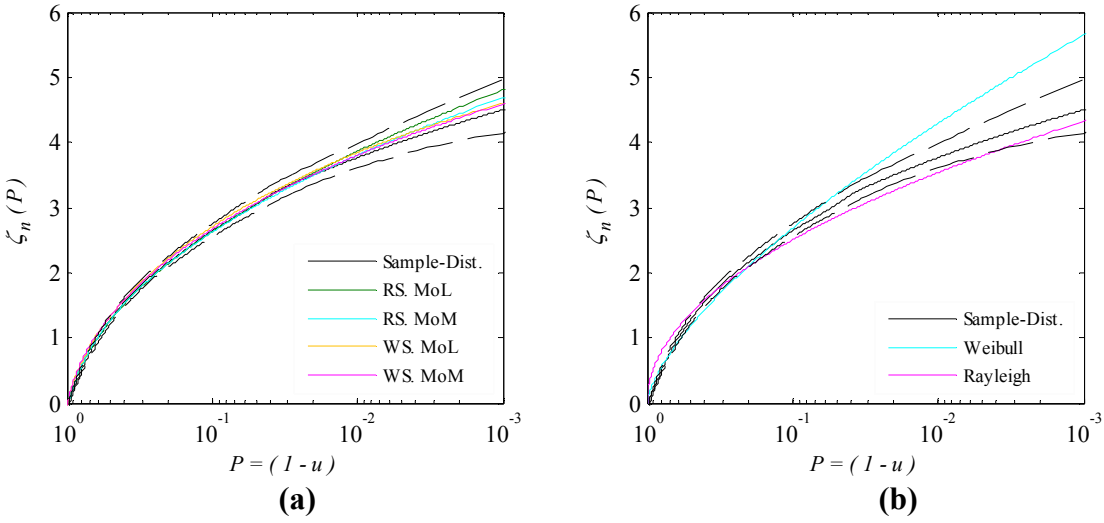


Figure 41 Exceedance probability distributions of normalized disturbed wave crest sample S2CA4 (a) semi-empirical models, (b) empirical models. Dashed lines represent the 95% confidence limits.

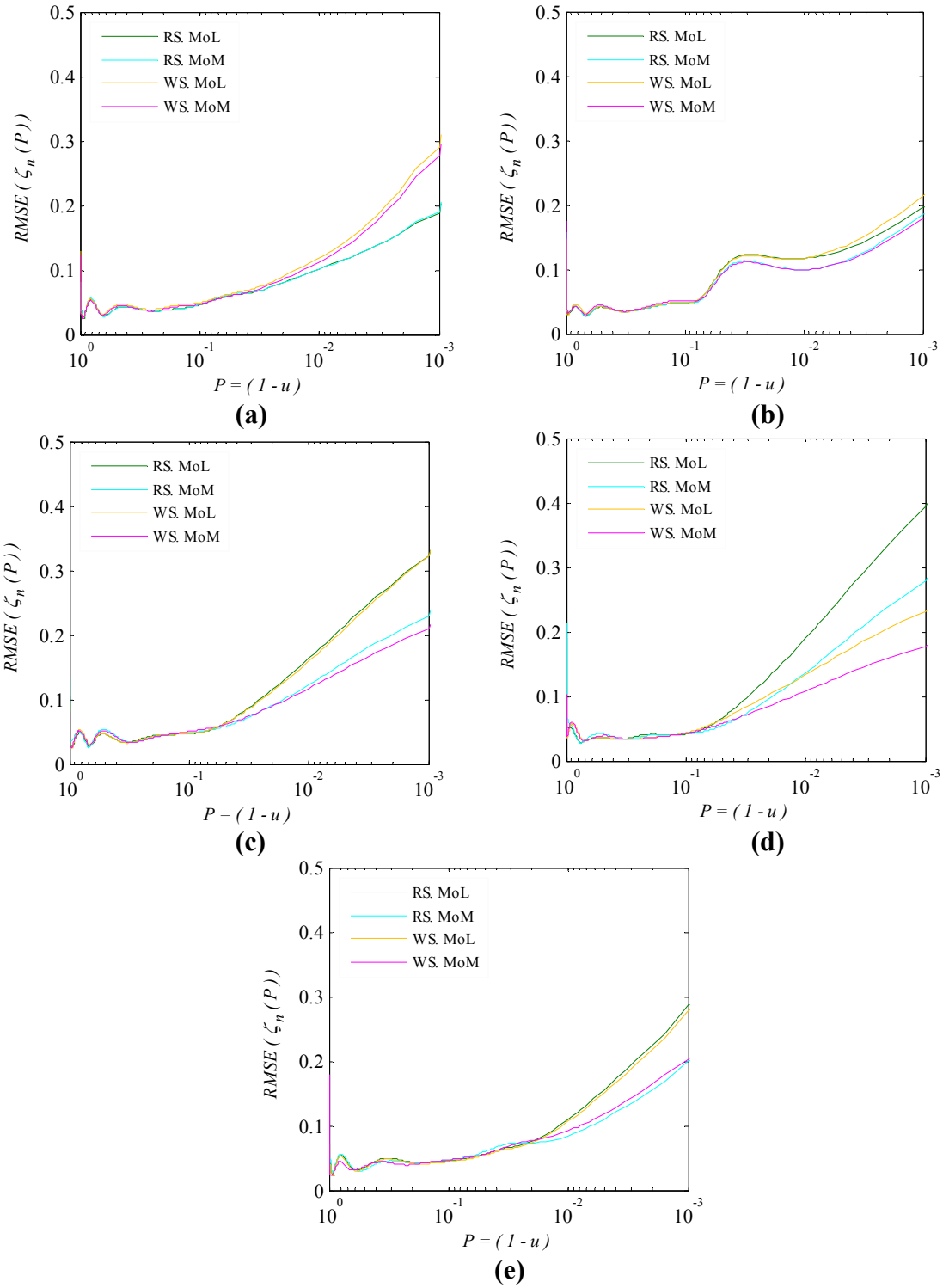


Figure 42 RMSE distribution of the semi-empirical models quantile estimates, (a) crest sample S1CA2, (b) crest sample S1CA3, (c) crest sample S1CA4, (d) crest sample S2CA3, (e) crest sample S2CA4.

The PDFs of wave crest maxima estimated from sample distribution and the semi-empirical models for the five disturbed wave crest samples are compared in Figure 43. The expected value estimates of these distributions are presented in Table 25. As shown in Figure 43, the difference between the model estimates and the sample distribution are more noticeable in these examples as compared to the similar distributions presented for undisturbed wave crests in Figure 36. The relatively small difference between the sample tail and the semi-empirical model estimates has been magnified in the PDF estimates of the wave crest maxima. It is important to notice that the crest maxima statistics vary significantly in the area close and underneath the mini-TLP. This can be recognized from the values shown in Table 25 for different samples. The maximum difference between the sample expected crest maxima and the semi-empirical models estimates is 8% of the sample value. Regarding the limited sample size and the uncertainty of the sample estimates, the semi-empirical models accuracy seems to be acceptable.

Table 25 Estimates of normalized expected crest maxima in 1,000 waves (disturbed wave samples).

Model	Disturbed wave crest samples				
	S1CA2	S1CA3	S1CA4	S2CA3	S2CA4
Sample Dist.	4.82	4.50	4.63	4.52	4.62
RS. MoL	4.76	4.72	4.40	4.29	5.00
RS. MoM	4.72	4.73	4.52	4.35	4.86
WS. MoL	4.54	4.62	4.36	4.40	4.72
WS. MoM	4.56	4.64	4.56	4.51	4.73

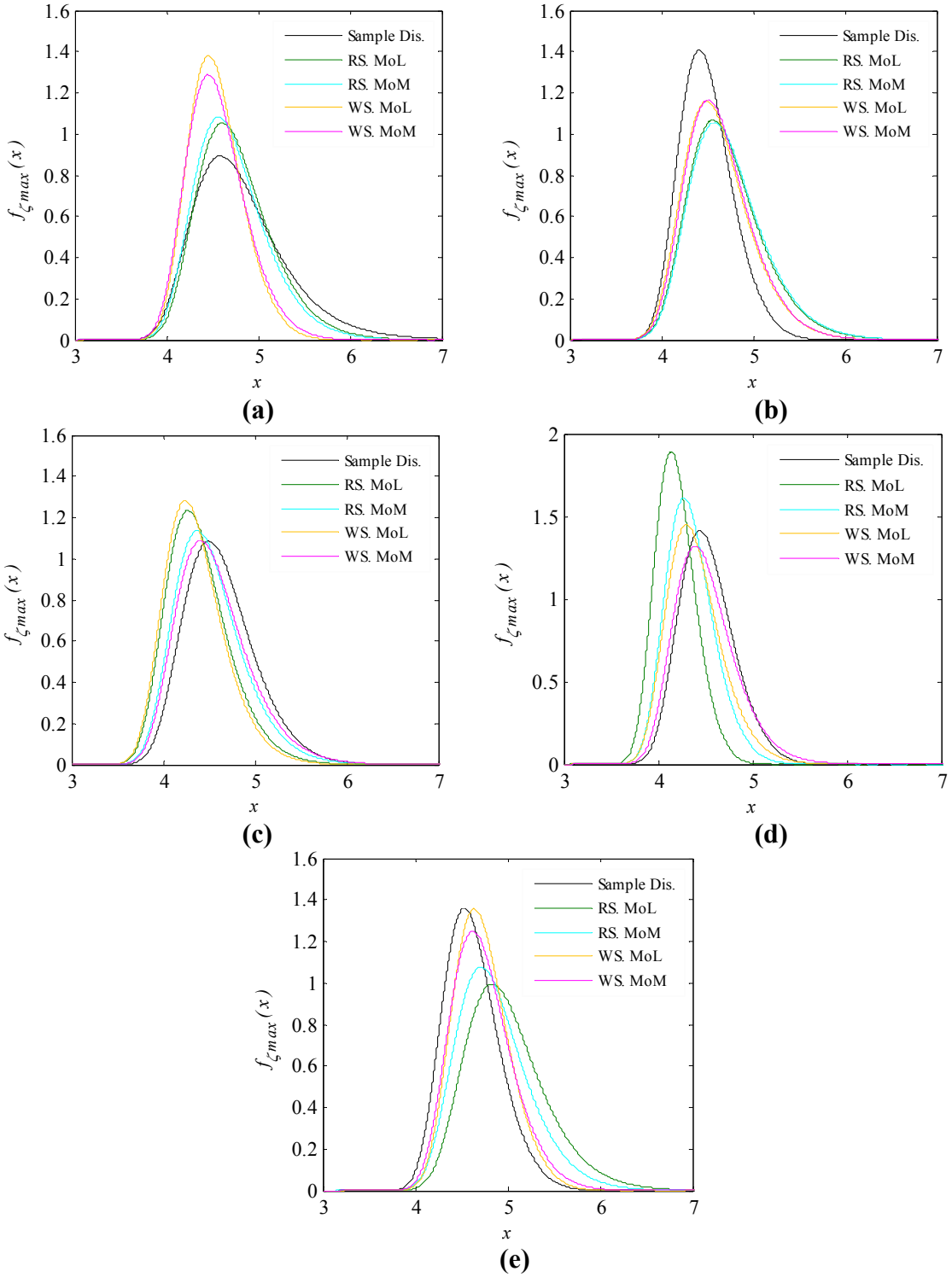


Figure 43 Probability density functions of disturbed normalized wave crest maxima in 1,000 waves, (a) crest sample S1CA2, (b) crest sample S1CA3, (c) crest sample S1CA4, (d) crest sample S2CA3, (e) crest sample S2CA4.

IV.5 Probability Distribution of Wave Run-up Over Offshore Structures Vertical Columns

IV.5.1 Introduction

Wave run-up over the vertical columns of offshore structures, similar to wave crests in the vicinity and underneath the structure, is an important design consideration for engineers tasked with deck elevation selection. Because of the complexity of the problem, there is no theoretical probability distribution to model the random variability of wave run-up in extreme environments which then leads engineers to employ empirical models.

Numerous research studies have been reported in the open literature that discuss the analytical solutions for linear and non-linear radiation-diffraction problems for regular waves. MacCamy and Fuchs [80] presented the seminal research study that addressed the analytical solution of the linear diffraction problem for a vertical cylinder extending from seabed and piercing the free surface. This was followed sometime later when the theoretical solution of the second-order diffraction problem subject to regular waves was derived and solved for finite and infinite water depth by Lighthill [81] and Molin [82]. The analytical and computational aspects of these non-linear theories have been widely pursued for a variety of offshore structures [83-87]. The prediction of wave run-up elevations on a vertical cylinder using diffraction theory has also been extensively investigated [88, 90]. Comparing the results of different models, Nielsen [91] showed that linear, second order and fully non-linear diffraction theories fail to estimate the resonance effects between the columns of offshore structures.

The application of empirical models to model the distribution of wave run-up due to random seas interacting with models of a Spar platform and a mini-TLP was investigated by Niedzwecki et al. [22]. Their study showed that wave run-up on circular columns such as these compliant platforms could be reasonably modeled using a two-parameter Weibull distribution. This finding was later confirmed using model test data from another model basin again in the study by Indrebo and Niedzwecki [23] that investigated the wave run-up on both a single rectangular column and an array of similar vertical columns. In these studies, it was also observed that a Rayleigh distribution model of the linear wave crest heights consistently underestimated the largest wave run-up observed.

A simplified theoretical model for estimating the non-linear wave run-up on vertical cylinders was proposed by Kriebel [17, 92]. It was based on three key assumptions: first, it was assumed that the incident wave field has narrow-band frequency spectrum; secondly, that the first and second order wave run-up are phase locked; and thirdly, that their maxima occur at the same time. Using this approach he was able to show that the simple superposition method was able to predict the run-up with accuracy comparable with more complete second order numerical calculations. This non-linear model was also used to estimate the probability distribution of wave run-up utilizing a methodology analogous to the one used by Tayfun [8] for estimating the probability distribution of non-linear wave crest height whose probability distribution was a function of two theoretically defined parameters.

In this study the three-parameter Rayleigh-Stokes (RS) and the four parameter Weibull-Stokes (WS) semi-empirical probability distribution models are utilized to estimate the probability distribution of wave run-up data. The model statistics estimated from method of L-moments (MoL) and method of moments (MoM) are evaluated. The RS semi-empirical model with the parameters estimated by MoL was previously applied by Izadparast and Niedzwecki [93, 94] to estimate the probability distribution of wave run-up over the columns of fixed and compliant platform configurations. The results of their analysis show promises in application of semi-empirical models for complex data sets. Here, following a similar methodology, the performance of the semi-empirical distributions is evaluated over sample data from the mini-TLP model test and the results are compared with empirical and theoretical models.

IV.5.2 Theoretical model

The total wave run-up on a vertical column may be considered to be the sum of three contributions, Kriebel [17], explicitly as,

$$R_n = R_1 + R_2 + \overline{R} \tag{4.34}$$

where, R_n is the total wave run-up and is the sum of R_1 the linear wave contribution, R_2 the second order wave contribution, and $\overline{R}$ the mean wave run-up on the cylinder. In Kriebel's model it is assumed that the first and second-order wave components are phase-locked and arrive at the cylinder simultaneously. Consequently, the total wave run-up is estimated as the superposition of the individual wave run-up components adjusted by the mean value. Although this simple transformation does not include all the

second-order wave run-up components, it accounts for the dominant run-up components, as noted, and the predictions using this simplified method were comparative to those estimated by a complete second order diffraction model [17]. A schematic introducing the main parameters involved in the run-up problem is presented in Figure 44.

In the Kriebel's model, the linear wave contribution is estimated using the linear diffraction theory and can be expressed as [95],

$$R_1 = \psi\left(\bar{k}\,r, \theta\right) a \tag{4.35}$$

and the amplification factor is defined as

$$\psi\left(\bar{k}\,r, \theta\right) = 2\left|\sum_{0}^{\infty} \frac{i\,\beta_m \cos\left(m\theta\right)}{\pi \bar{k}\,r\, H_m'^{(1)}\left(\bar{k}\,r\right)}\right| \tag{4.36}$$

where, r is the radius of the vertical column, θ is the angle measured from the rear center of the cylinder, both introduced in Figure 44. In these equations, $\bar{k}\,r$ is the scatter parameter, r is the radius of the vertical column, the angle θ is measured from the rear center of the cylinder, $H_m^{(1)}$ is a Hankel function of the first kind, the prime indicates differentiation with respect to the argument, and $\beta_0 = 1,\ \ \beta_m = 2i^m$ for $m \geq 1$. The variation of the linear amplification factor ψ in terms of the scatter parameter $k_0 r$ and angle θ is shown in Figure 45. As shown in this figure, the incident waves remain intact when the scatter parameter is small while the incident wave crest height almost doubles in front of a large column. Figure 45 clearly shows the re-elevation of wave run-up on the back side of the column at $\theta = 0$ after a dramatic elevation drop on the sides of the column.

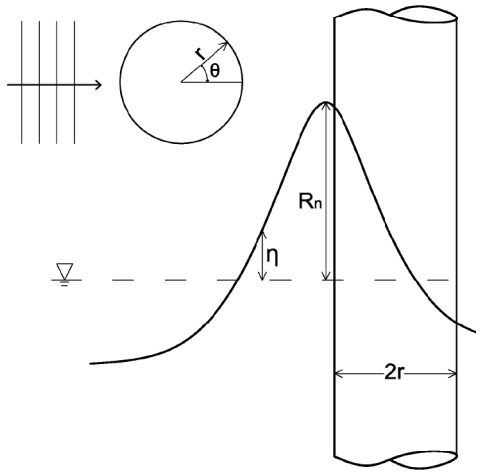


Figure 44 Wave run-up over a vertical column.

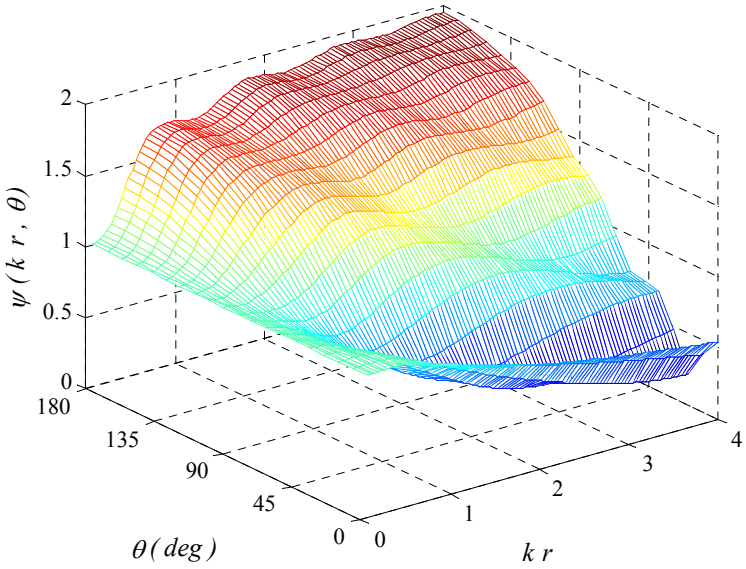


Figure 45 Run-up linear amplification factor.

Kriebel [17] assumed that the second-order harmonic travels two times as fast as the first harmonic and has a wave number of $2\bar{k}$. R_2 is then approximated from

modification of the second-order Stokes wave amplitude, $\bar{k}a^2/2$, utilizing the linear diffraction amplification factor ψ with scatter parameter of $2\bar{k}r$, specifically,

$$R_2 = \psi\left(2\bar{k}r, \theta\right)\frac{\bar{k}a^2}{2} \tag{4.37}$$

An estimate for the mean run-up $\bar{R}$ in deep water consistent to second-order may be obtained from [92],

$$\bar{R} = \frac{\bar{k}a^2}{4}\psi^2\left(\bar{k}r, \theta\right) \tag{4.38}$$

Upon substituting Eq.(4.35), (4.37), and (4.38) in Eq. (4.34), the expression for non-linear run-up R_n may be given in the form,

$$\begin{aligned} R_n &= a\,\psi\left(\bar{k}r, \theta\right) + \frac{\bar{k}a^2}{2}\psi\left(2\bar{k}r, \theta\right) + \frac{\bar{k}a^2}{4}\psi^2\left(\bar{k}r, \theta\right) \\ &= a\,\psi\left(\bar{k}r, \theta\right) + a^2\frac{\bar{k}}{2}\left[\psi\left(2\bar{k}r, \theta\right) + \frac{1}{2}\psi^2\left(\bar{k}r, \theta\right)\right] \end{aligned} \tag{4.39}$$

Normalizing Eq. (4.39) with respect to the standard deviation of incident wave surface elevation σ_η, one obtains the compact expression for the non-linear run-up as,

$$\zeta_n = \alpha\zeta + \beta\zeta^2 + \gamma \tag{4.40}$$

where, $\zeta_n = R_n/\sigma_\eta$ and $\zeta = a/\sigma_\eta$ are non-dimensional wave run-up and non-dimensional incident wave crest height, respectively. The transformation equation introduced in Eq. (4.40) is the same as the perturbation expansion defined in Eq. (2.3). Modeling the linear wave crests with the Rayleigh distribution with parameter $R = 1.0$, Kriebel [17] derived a probability distribution for wave run-up over offshore vertical

columns. Kriebel's theoretical probability distribution has a structural form similar to the Rayleigh-Stokes distribution Eq. (2.43) and the parameters are approximated as,

$$\begin{aligned}\alpha &= \psi\left(\bar{k}\, r, \theta\right), \\ \beta &= \left[\psi\left(2\bar{k}\, r, \theta\right) + \frac{1}{2}\,\psi^2\left(\bar{k}\, r, \theta\right)\right]\frac{\bar{k}\sigma_\eta}{2}, \\ \gamma &= 0.\end{aligned} \tag{4.41}$$

It is straightforward to utilize Kriebel's approximation, as one only needs to estimate the deepwater mean frequency and the standard deviation from the significant wave height.

The other theoretical model that has been widely used to approximate the probability distribution of wave run-up over vertical columns is the Rayleigh distribution of linear narrow-banded process. Although it has been shown by research studies that the model significantly understates the run-up distribution, the theoretical Rayleigh distribution has been widely used in the practical engineering problems because of its simplicity.

IV.5.3 Empirical probability distributions

The empirical distributions used here to estimate the probability distribution of wave run-up data sets are the ones utilized for disturbed and undisturbed wave crests, i.e. empirical Rayleigh and two-parameter Weibull distributions. The details of these models are discussed in section IV.4.3.

IV.5.4 Sample data sets

The wave run-up data sets used here are obtained from the mini-TLP model test described in section IV.2. Here, the measurements at run-up probes R1 and R2 (see Figure 19) of the fixed hull configuration are analyzed. The random long-crested incident waves in the studied cases travel in quartering sea orientation of the model test and correspond to the seastates S-1 and S-2 introduced in Table 14. As mentioned earlier, the unmanned mini-TLP was initially designed for deployment off the coast of West Africa but for the model test program the deck elevation was increased by 5m (prototype scale) in order to accommodate more severe seastates, i.e. S-2. During these tests wave over topping and wave impact on the deck structure were observed. This lead to a discrepancy since the actual wave heights were cut-off as they exceeded the measurement capability of the wave gages that were applied to surfaces of the vertical TLP columns. The associated data points were removed from the data being analyzed, but the probability distributions were calculated recognizing this issue.

To be consistent with other research studies, the wave run-up is defined as the maximum wave elevation in between each two consecutive zero up-crossings. Similar to wave crest samples, each wave run-up sample is developed from zero-crossing analysis of a 3-hour run-up elevation time series.

IV.5.5 Analysis and results

The first four L-moments and moments of the wave run-up samples are presented in Table 26. In this table, a sample name indicates the seastate and the location at which the sample was measured, e.g. the sample S1R1 represents the run-up sample measured

during the seastate S-1 at run-up probe R1. In estimation of these statistics the samples are normalized by the first-order standard deviation of incident waves σ_η estimated from its relationship with the significant wave height $\sigma_\eta = H_s/4$. In Table 26 the sample estimates are compared with the statistics of the theoretical Rayleigh distribution (second column) and Kriebel's theoretical model (KR). As shown in Table 26, the samples have significantly large first and second moments (L-moments) and the situation is more critical in samples measured during the second seastate, i.e. S-2. This has not been well approximated by the theoretical models as both theoretical distributions have considerably smaller first and second moments. As expected Kriebel's model better approximates the sample statistics as compared to the theoretical Rayleigh distribution. In the most studied cases, the third moment is overestimated by Kriebel's model. Another major different between the sample estimates and the model statistics happens in the fourth moment (L-moment). Despite the sample S1R1, the other three samples have negative kurtosis while the theoretical models, especially Kriebel's model, have significantly large positive kurtosis. From these, one can conclude that the theoretical models are unsuccessful in representing the basic sample statistics and consequently are not expected to be able to approximate the sample distributions accurately.

Table 26 L-moments and moments of normalized wave run-up.

Estimate	Rayleigh Distribution	Wave run-up samples							
		S1R1		S1R2		S2R1		S2R2	
		Sample	KR	Sample	KR	Sample	KR	Sample	KR
$l_1(\zeta_n)$	1.253	1.450	1.297	1.369	1.286	1.979	1.439	1.562	1.406
$l_2(\zeta_n)$	0.367	0.558	0.387	0.515	0.384	0.746	0.456	0.581	0.444
$\hat{\tau}_3(\zeta_n)$	0.114	0.133	0.124	0.115	0.124	0.127	0.154	0.112	0.152
$\hat{\tau}_4(\zeta_n)$	0.105	0.083	0.108	0.078	0.108	0.067	0.117	0.085	0.116
$\hat{\mu}_1(\zeta_n)$	1.253	1.450	1.297	1.369	1.286	1.979	1.439	1.562	1.406
$\hat{\mu}_2(\zeta_n)$	0.429	0.998	0.482	0.835	0.473	1.734	0.683	1.063	0.645
$\hat{s}_{\zeta_n}$	0.631	0.750	0.695	0.587	0.693	0.534	0.882	0.585	0.868
$\hat{K}_{\zeta_n}$	0.245	0.525	0.404	-0.096	0.399	-0.532	0.930	-0.041	0.889

Utilizing the sample statistics given in Table 26, the empirical estimates of the RS and WS model parameters are given in Table 27. In estimation of these parameters, without loss of generality, it is assumed that the RS model parameter $R = 1.0$ and the WS model parameter $\vartheta = \sqrt{2}$. The estimates of the theoretical Kriebel model parameters are given in Table 28. It has been seen that the results of the two parameter estimation methods, i.e. MoM and MoL, are fairly close in both RS and WS models while the difference between the parameter estimates of different models is more recognizable. As shown in Table 27 the semi-empirical models have captured considerably large first-order contribution α which is primarily caused by the samples large second moment (L-moment). The theoretical estimates of the parameter α, shown in Table 28, are considerably smaller than the empirical ones. Except for the sample R1S1, the second-order terms have negative contribution to the semi-empirical distributions. This is

recognized by negative values of parameter β in Table 27. It has been observed that the WS model has constantly smaller parameter β values as compared to the RS model. The theoretical estimate of parameter β is always positive and the parameter has larger values for the measurements during the more energetic seastate S-2. Except for the sample S1R1, the estimates of the WS shape parameter κ are significantly different from the RS pre-assumed shape parameter of $\kappa = 2.0$. The semi-empirical models have estimated negative shifting parameter for the samples being studied here, which indicates that the run-up distributions are shifted towards smaller run-up. Note that the shifting parameter γ is assumed to be zero in the theoretical model. As shown in Table 27, the RS model constantly has smaller shifting parameter γ than the WS model. The negative shifting in the distribution causes non-zero probability for negative wave run-up which is not theoretically justified. To solve this issue, it is assumed that $P(\zeta_n = 0) = P(\zeta_n \le 0)$.

Table 27 Empirically estimated parameters of Rayleigh-Stokes and Weibull-Stokes models for the normalized wave run-up samples.

Parameter	Model	Method	Wave run-up samples			
			S1R1	S1R2	S2R1	S2R2
$\hat{\alpha}$	RS	MoL	1.388	1.398	1.907	1.598
	RS	MoM	1.381	1.444	2.169	1.632
	WS	MoL	1.417	1.419	1.915	1.627
	WS	MoM	1.397	1.387	1.919	1.616
$\hat{\beta}$	RS	MoL	0.048	0.002	0.046	-0.006
	RS	MoM	0.049	-0.017	-0.055	-0.020
	WS	MoL	0.019	-0.082	-0.161	-0.108
	WS	MoM	0.024	-0.080	-0.163	-0.091
$\hat{\kappa}$	RS	MoL	NA	NA	NA	NA
	RS	MoM	NA	NA	NA	NA
	WS	MoL	1.941	1.643	1.469	1.612
	WS	MoM	1.921	1.608	1.446	1.673
$\hat{\gamma}$	RS	MoL	-0.386	-0.387	-0.504	-0.429
	RS	MoM	-0.379	-0.407	-0.631	-0.443
	WS	MoL	-0.342	-0.220	-0.105	-0.234
	WS	MoM	-0.352	-0.231	-0.090	-0.280

Table 28 Theoretical estimates of the Rayleigh-Stokes model parameters for the normalized run-up samples.

Parameter	Model	Wave run-up samples			
		S1R1	S1R2	S2R1	S2R2
$\hat{\alpha}$	KR	1.005	0.9980	1.014	0.998
$\hat{\beta}$	KR	0.019	0.0178	0.084	0.077
$\hat{\gamma}$	KR	0.000	0.000	0.000	0.000

The exceedance probability distributions of wave run-samples are compared with estimates of semi-empirical, empirical, and theoretical models for the run-up samples S1R1, S1R2, S2R1, and S2R2 in Figure 46, Figure 47, Figure 48, and Figure 49, respectively. In these figures, the sample distribution is estimated from the semi-parametric approach and the dashed lines indicate the 95-percent confidence intervals of the sample distribution estimated utilizing 10,000 bootstrap samples. A common pattern in all the four samples studied here is that the sample exceedance probability distribution starts with a relatively steep slope mainly caused by the large second moment. The sample distributions of these four samples are significantly different in the tail part. As shown in Figure 46, the S1R1 sample distribution has a heavy tail while the sample measured on the back side of the column during the same seastate, S1R2, has a flatter tail (see Figure 47). The tail part of the S2R1 sample distribution (see Figure 48) shows a sudden slope change resulting in a flat tail. This, however, may be caused by the measurement limitation described earlier. The distribution slope change is milder in case of the sample S2R2 resulting in a heavier tail as compared to the tail distribution of the S2R1 (see Figure 49).

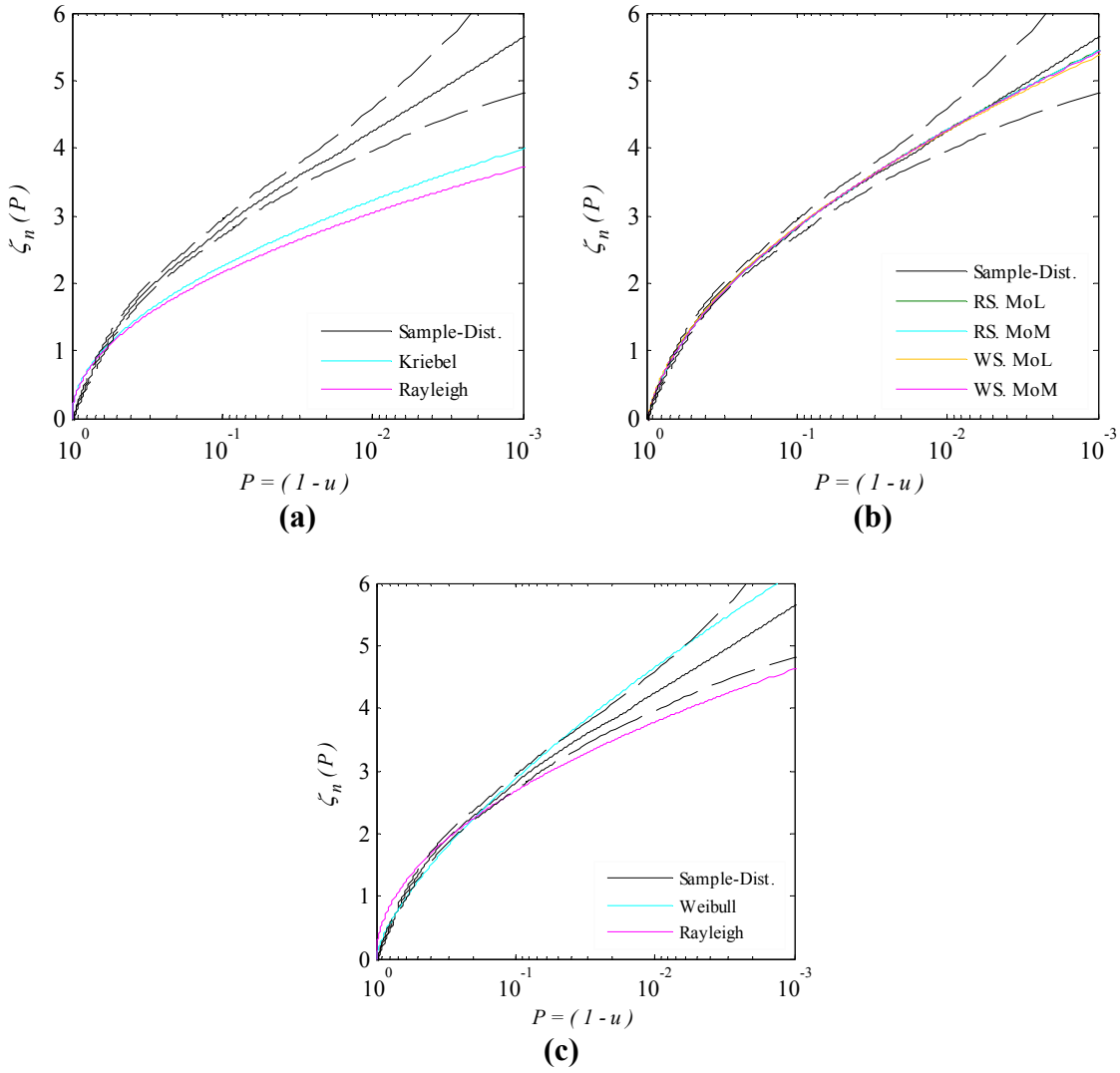


Figure 46 Exceedance probability distributions of wave run-up sample S1R1 (a) theoretical models, (b) semi-empirical models, (c) empirical models. Dashed lines represent the 95% confidence limits.

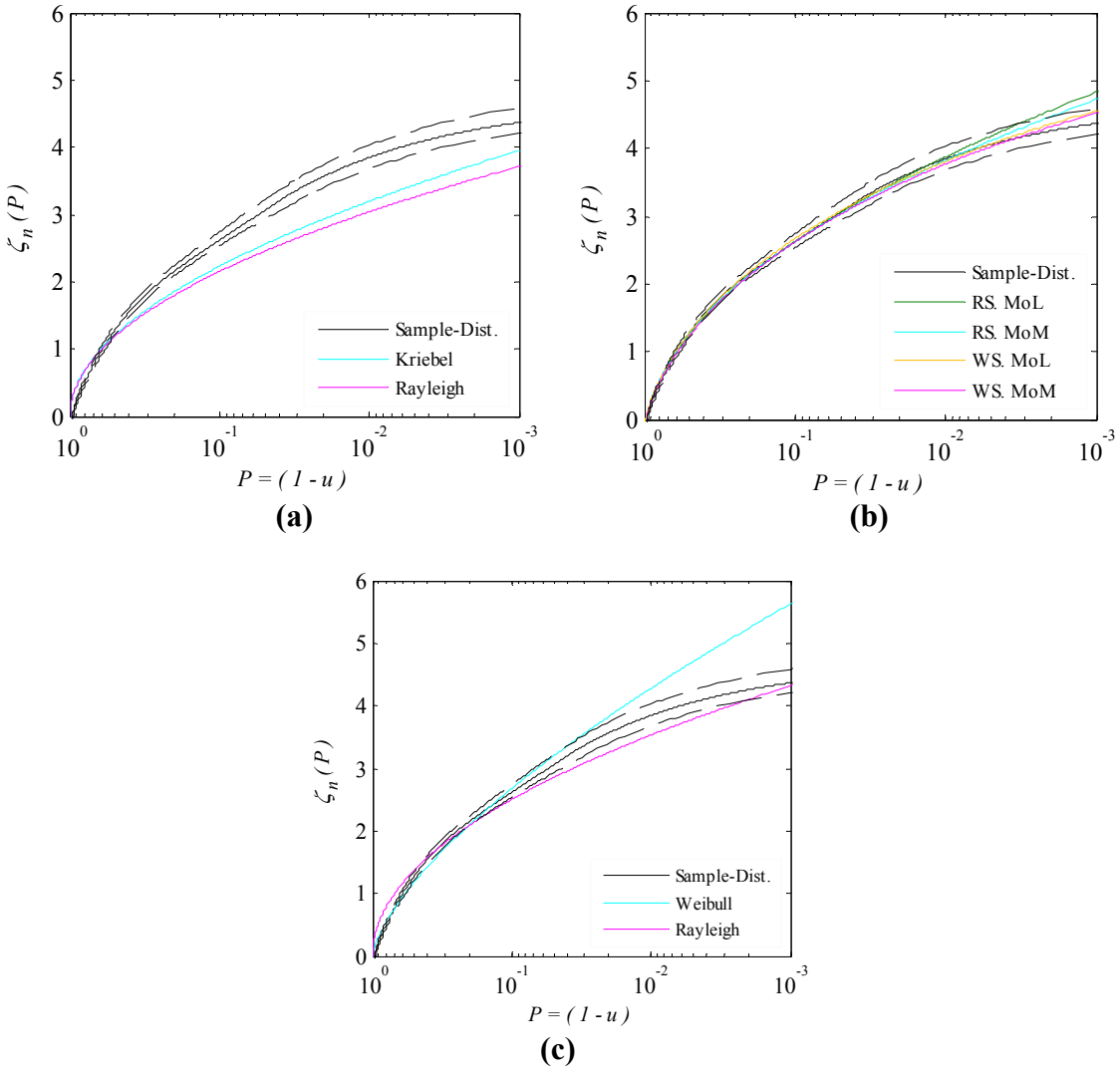


Figure 47 Exceedance probability distributions of wave run-up sample S1R2 (a) theoretical models, (b) semi-empirical models, (c) empirical models. Dashed lines represent the 95% confidence limits.

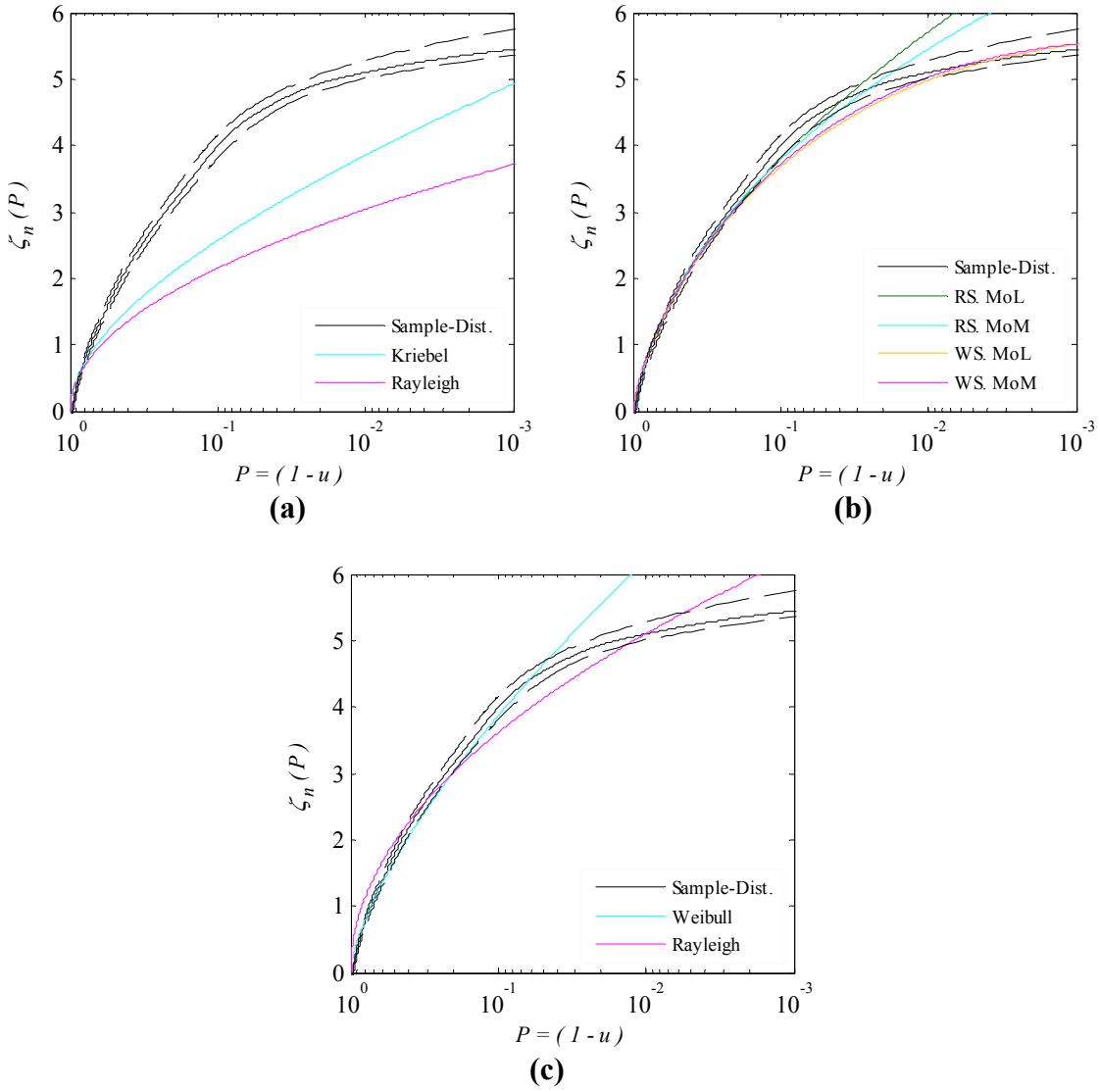


Figure 48 Exceedance probability distributions of wave run-up sample S2R1 (a) theoretical models, (b) semi-empirical models, (c) empirical models. Dashed lines represent the 95% confidence limits.

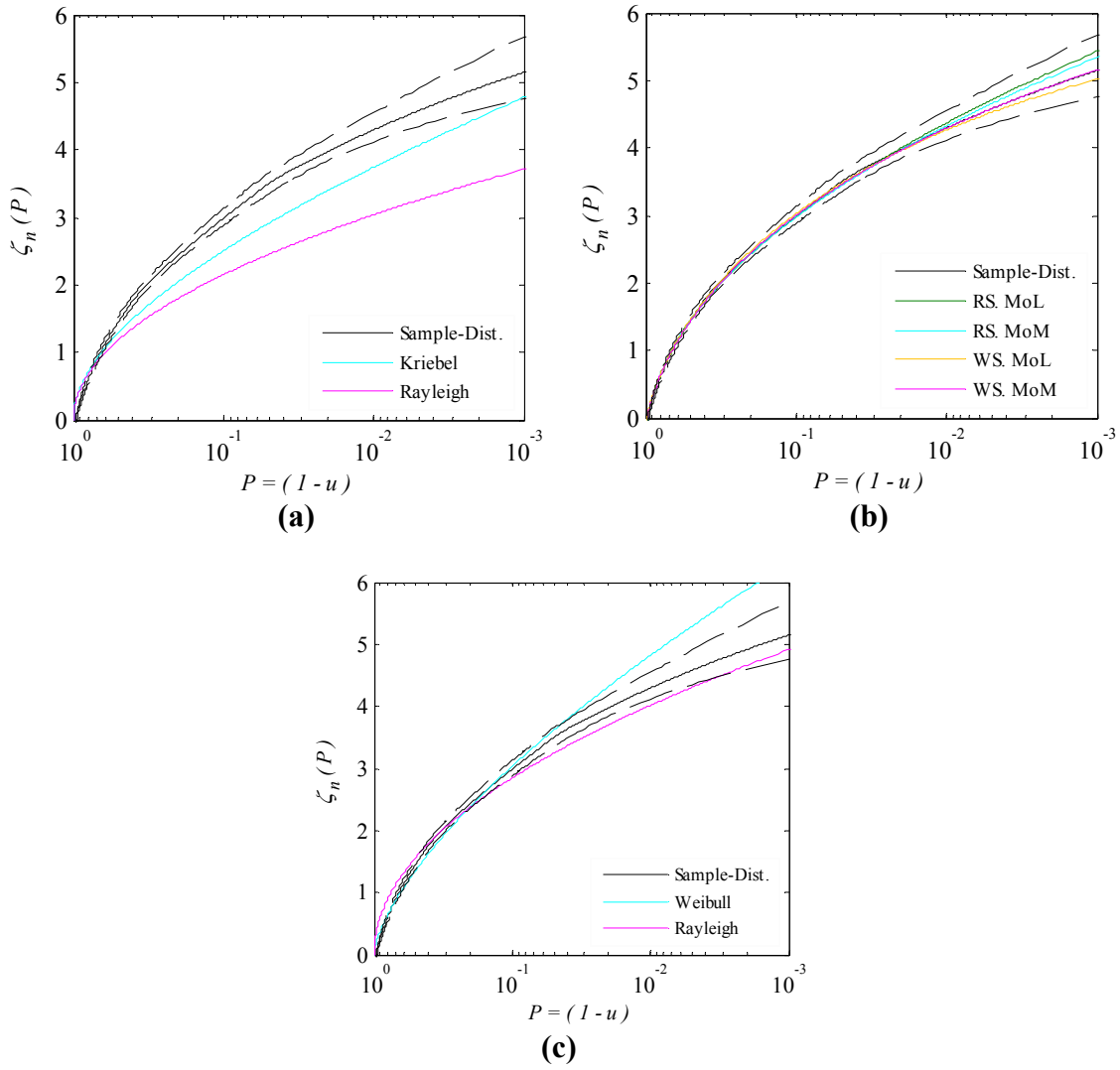


Figure 49 Exceedance probability distributions of wave run-up sample S2R2 (a) theoretical models, (b) semi-empirical models, (c) empirical models. Dashed lines represent the 95% confidence limits.

As shown in Figure 46 to Figure 49 the theoretical models are not successful in representing the probability distribution of sample data. Both theoretical Rayleigh and Kriebel distributions significantly underestimate the probability distribution of wave run-up. Among the empirical models, the two-parameter Weibull distribution performs considerably better than the empirical Rayleigh distribution. The empirical Rayleigh model fails to estimate the probability distribution of small and large wave run-ups. In the cases studied here, the empirical Weibull model estimates the major part of the distribution with $P > 0.1$ reasonably accurately; however, the model tends to overestimate the tail part of the distributions.

Regarding the distributions shown in Figure 46 to Figure 49, it can be concluded that the semi-empirical models are robust in estimating the probability distribution of wave run-up. The distribution estimates of the four semi-empirical models being studied in this study, i.e. RS MoL, RS MoM, WS MoL, and WS MoM are pretty close for the major part of the distribution. The difference between the model distributions happens on the tail part of the distributions. Similarly to what was observed in the previous section for the wave crest distributions, the sample tail may not follow the same pattern as the majority of the distribution. The sample distributions routinely show a sudden slope change in the tail part of the distribution which may be caused by an energy loss mechanism, e.g. wave breaking, or an increase in the contribution of the negative second-order terms. In the run-up examples studied here, a variety of these effects are observed. The tail distribution of the sample S1R1 (see Figure 46) follows the same pattern as the rest of the measurements and no sensible slope change has observed in the

probability distribution. Consequently, the semi-empirical models successfully estimate the probability distribution of data and the estimates of different models even on the tail part of the distributions are considerably close. The tail distributions of the measurements on the back side of the column at R2 in both seastates have a milder slope than the rest of the distribution which results in a relatively flat tail (see Figure 47 and Figure 49). The tail parts of different semi-empirical models are slightly different for these cases. As shown in Figure 47 (b) and Figure 49 (b) the WS model because of utilization of the fourth moment (L-moment) is more flexible in capturing the slope change in the tail part of the sample distributions as compared to the RS model. In these cases the RS model tends to overestimate the tail distributions. The run-up sample S2R1 has a complex distribution (see Figure 48) which starts with a steep slope and experiences a dramatic slope change and a flat tail. The sample tail in this case is contaminated by the measurement limitation and contains a high level of uncertainty. Note that the confidence intervals, shown in these figures, are basically indications of the variability of the bootstrap samples, and should not be confused with the uncertainty of the original measurements. Interestingly, as shown in Figure 48 (b), the WS model is able to capture the slope change in the sample distribution while the RS model is affected by the initial steep slope and overestimate the tail distribution.

The RMSE in quantile estimates of the semi-empirical models for the four run-up samples are presented in Figure 50. A similar discussion as made in above can be made here. Despite for the run-up sample S2R1 where the sample quality is questionable, the error in the semi-empirical model estimates remains fairly small for other examples.

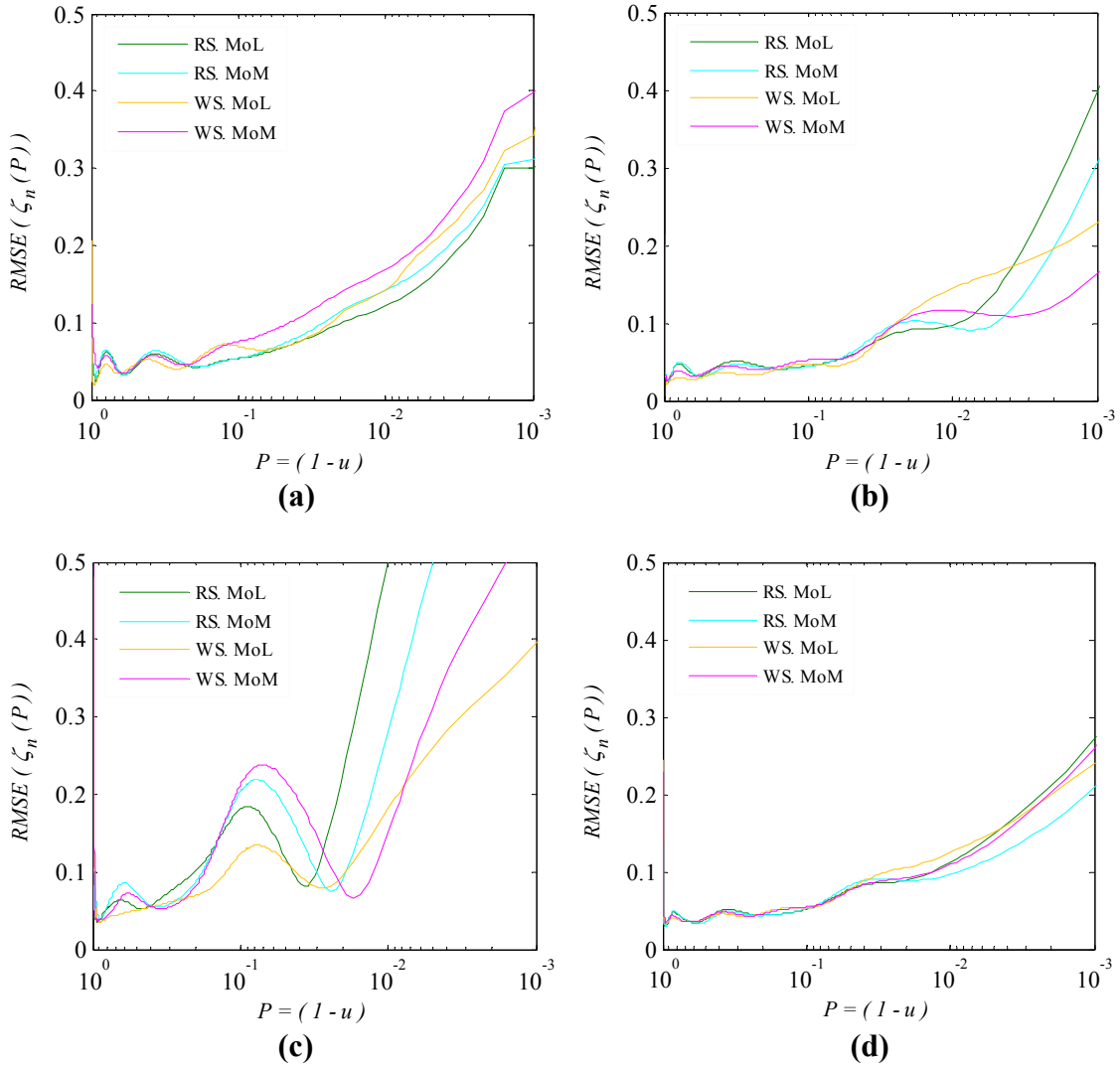


Figure 50 RMSE distributions of the semi-empirical models quantile estimates, (a) run-up sample S1R1, (b) run-up sample S1R2, (c) run-up sample S2R1, (d) run-up sample S2R2.

The PDFs of run-up maxima in $N = 1,000$ waves for each of the four examples are presented in Figure 51 and the estimates of the expected run-up maxima are compared in Table 29. The statistics of the run-up maxima are estimated utilizing the extreme value theory described in section II.6. As shown here, the WS model provides more realistic estimates of the extreme statistics. It is interesting to note that both RS and WS distributions have an upper limit of $x_{upper} = \gamma - \frac{\alpha^2}{4\beta}$ for $\beta < 0$. In most cases with negative β the ratio $-\frac{\alpha^2}{4\beta}$ is considerably large and therefore the upper limit is out of the practical range of wave run-up. However, in case of the sample S2R1, the WS models have considerably large absolute β parameter resulting in an upper limit of approximately $x_{upper} \approx 5.6$. This has caused the special shape of the run-up maxima PDF estimated by WS models shown in Figure 51 (c). As shown in this figure, the PDF has a sharp peak and ends at the upper limit which actually follows a similar pattern as the sample distribution.

Table 29 Estimates of normalized expected run-up maxima in 1,000 waves.

Model	Wave run-up samples			
	S1R1	S1R2	S2R1	S2R2
Sample Dist.	5.85	4.51	5.48	5.20
RS. MoL	5.68	5.03	7.57	5.64
RS. MoM	5.67	4.91	6.95	5.55
WS. MoL	5.60	4.67	5.56	5.12
WS. MoM	5.65	4.63	5.55	5.29

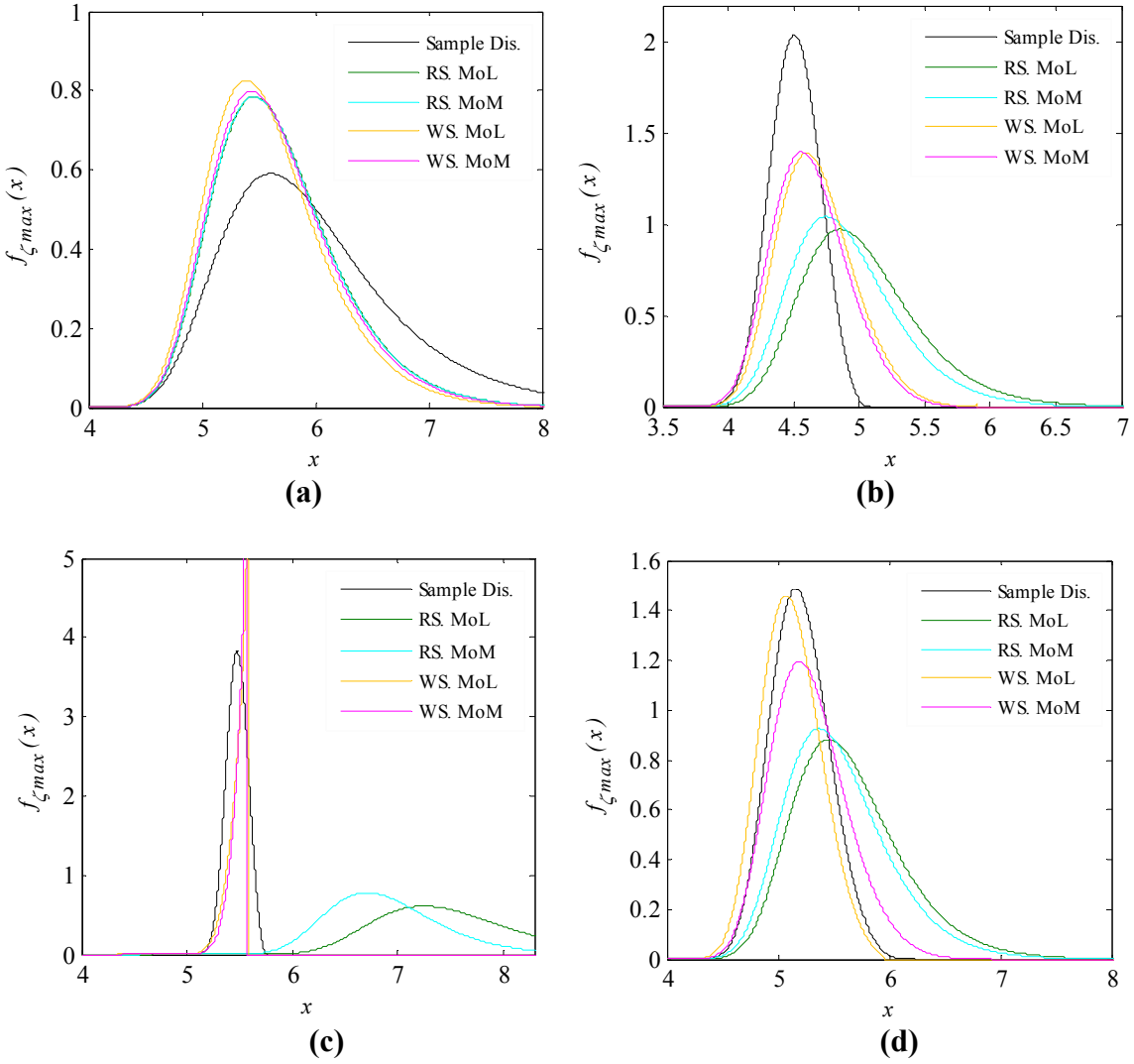


Figure 51 Probability density functions of normalized wave run-up maxima in 1,000 waves, (a) run-up sample S1R1, (b) run-up sample S1R2, (c) run-up sample S2R1, (d) run-up sample S2R2.

IV.6 Probability Distribution of Wave Power

IV.6.1 Introduction

The process of developing technologies for the efficient extraction of energy from ocean waves provides a range of technical and practical challenges. One of the most critical challenges initially, is the need to accurately characterize the random behavior of ocean waves and to design wave energy conversion devices that can function optimally for a range of wave frequencies at various offshore sites. The first step is to develop reliable estimates of the short and long-term characterizations of the wave fields of interest. The yearly mean wave power is commonly considered to be the characteristic value used for design purposes, see for example [96-100]. This long-term value can be obtained by estimating the mean wave power in a sea-state and the associated probabilities obtained from sea-states scatter diagram. In this particular approach the wave power is basically characterized by a single value and consequently the uncertainty of the process cannot be adequately captured.

Recently, Myrhaug et al. [24, 25] studied the probability distribution of wave power in a sea-state using an empirical bi-variate distribution of wave height and period for waves on the Norwegian continental shelf. In their study the random variable transformation was utilized in order to obtain the bi-variate probability distributions of wave power and the associated sea-state parameters of wave height and period. Izadparast and Niedzwecki [101] pursued a methodology similar to Myrhaug et al. [25] but the bi-variate probability distributions of wave power, wave height and wave period were derived using the theoretical joint probability distributions [60, 61]. In the study by

Izadparast and Niedzwecki [101], the dimensionless asymptotic forms of these distributions for both the shallow and deepwater limits are developed. Additionally, the distributions were simplified for waves with narrow-banded spectra. The statistics of these theoretical models were compared with those obtained from a wave-by-wave analysis of waves generated from a JONSWAP spectrum for sea-states of varying severity. The severity of the wave conditions is investigated based upon the variation of the JONSWAP model peakedness parameter. It was shown that the theoretical models are successful in representing the probability distribution of wave power samples generated from wave-by-wave analysis of simulated wave records. In this study a special case of Rayleigh-Stokes (RS) semi-empirical model with no linear term is utilized to estimate the probability distribution of wave power data. Eventually, the statistics of the semi-empirical model are compared with those obtained from the theoretical models and wave-by-wave analysis of the simulated wave records.

IV.6.2 Mathematical background

The wave power per unit crest length P of regular seas as developed from linear wave theory is well-known and can be expressed as the product of the wave group velocity c_g and the total average wave energy per unit surface area E, i.e. [42],

$$P = C_g E \tag{4.42}$$

For the deepwater limit, indicated by the subscript d, with depth d to wave number λ ratio of $d/\lambda > 1/2$, the wave power can be expressed as,

$$P_d = \frac{\rho g^2}{32\pi} T H^2 \tag{4.43}$$

where, g is the gravitational acceleration, ρ is the fluid density, T is the wave period, and H is the wave height. For the shallow water limit, denoted by the subscript s, with $d/\lambda < 1/20$, the wave power can be expressed as,

$$P_s = \frac{\rho g^{3/2}}{8} d^{1/2} H^2 \tag{4.44}$$

In this limit, the wave power varies with the square-root of water depth d, rather than the wave period as in the deepwater limit. Normalizing Eq. (4.43) with a characteristic wave power $\hat{P}_d = \frac{\rho g^2}{64\pi} T_p H_s^2$, the dimensionless form of wave power in deepwater p_d is obtained as,

$$p_d = \frac{P_d}{\hat{P}_d} = \beta_d \zeta^2 \tag{4.45}$$

where $\zeta = H/\sigma_\eta$, is the dimensionless wave height, $\sigma_\eta = m_0^{1/2} = H_s/4$ is the standard deviation of surface wave elevation, and the model parameter $\beta_d = \frac{1}{8}\frac{T}{T_p}$. The model parameter in this case varies as a function of wave power. For waves with narrow-banded spectrum, the wave energy is concentrated in the vicinity of the peak period and consequently the model parameter $\beta \to 1/8$. Similarly, the dimensionless form of wave power in shallow water is obtained from normalizing Eq. (4.44) with the characteristic wave power in shallow water $\hat{P}_s = \frac{\rho g^{3/2}}{16} d^{1/2} H_s^2$, particularly,

$$p_d = \frac{P_s}{\hat{P}_s} = \beta_s \zeta^2 \tag{4.46}$$

where the model parameter in this case has a constant value of $\beta_s = 1/8$.

IV.6.3 Theoretical probability distributions

The joint distribution of the dimensionless wave height $\zeta = H/\sigma_\eta$ and wave period $\tau = T/\overline{T}$ for a given sea-state S can be expressed as [60, 61],

$$f_{\zeta,\tau}(x, t \mid S) = \frac{1}{8\nu_s\sqrt{2\pi}} L(\nu_s) \left(\frac{x}{t}\right)^2 \exp\left[-\frac{x^2}{8}\left(1+\left(1-\frac{1}{t}\right)^2 \frac{1}{\nu_s{}^2}\right)\right] \tag{4.47}$$
$$0 \le x < \infty,\ 0 < t < \infty$$

For convenience, the combination of parameters $L(\nu_s) = 2\left[1+\left(1+\nu_s{}^2\right)^{-1/2}\right]^{-1}$ is utilized. From this, the marginal probability distribution of normalized wave height $f_\zeta(x \mid S)$ in the sea-state S is obtained in the form of,

$$f_\zeta(x|S) = \int_0^\infty f_{\zeta,\tau}(x,t|S)\, dt = \frac{x}{4}\exp\left(-x^2/8\right) L(\nu_s)\, \Phi\left(\frac{x}{2\nu_s}\right) \tag{4.48}$$

where, the normal distribution function is

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-u^2/2\right) du \tag{4.49}$$

In case of narrow-banded waves, the probability distribution of normalized wave height asymptotes to Rayleigh distribution, Eq. (2.42), with parameter $R = 4.0$. Utilizing the joint probability distribution in Eq. (4.47) and the relationship between the wave

power and wave height, Eq. (4.45), the joint probability distribution of $f_{\zeta,p_d}\left(x,\not{\rho}\middle|S\right)$ is estimated as [101],

$$f_{\zeta,p_d}\left(x,\not{\rho}\middle|S\right)=\frac{\bar{T}\,\zeta^4}{64T_p\sqrt{2\pi}\,\nu_s\not{\rho}^2}L(\nu_s)\exp\left[\frac{-x^2}{8}\left(1+\left(1-\frac{\bar{T}x^2}{8T_p\not{\rho}}\right)^2\frac{1}{\nu_s{}^2}\right)\right] \quad (4.50)$$

Knowing the joint probability distribution of wave power and wave height, the short-term probability distribution of wave power $f_{p_d}\left(\not{\rho}\middle|S\right)$ can be obtained from the integration $f_{p_d}\left(\not{\rho}\middle|S\right)=\int_0^\infty f_{\zeta,p_d}\left(x,\not{\rho}\middle|S\right)\partial x$. In this distribution the random variability of both wave height and wave period along with the linear correlation between these random variables are considered. The situation is more straight forward for shallow water condition since the model parameter is not a function of wave period and the probability distribution of wave power can be estimated utilizing the relationship Eq. (4.46) in conjunction with the probability distribution in Eq. (4.48), specifically,

$$f_{p_s}\left(\not{\rho}\middle|S\right)=\left(1+\frac{\nu_s{}^2}{4}\right)\exp\left(-\not{\rho}\right)\Phi\left(\frac{\sqrt{2\not{\rho}}}{\nu_s}\right) \quad (4.51)$$

Assuming that waves have narrow-banded energy spectrum with $\nu_s\to 0$, the marginal probability distribution of p_d and p_s are derived in the form of,

$$f_{p_d}\left(\not{\rho}\middle|S\right)=f_{p_s}\left(\not{\rho}\middle|S\right)=\exp\left(-\not{\rho}\right) \quad (4.52)$$

which is the well-known exponential distribution with the expected value and standard deviation of unity, i.e. $E\left(p_d\right)=E\left(p_s\right)=1.0$ and $\sigma_{p_d}=\sigma_{p_s}=1$. This approximation for

deepwater condition assumes that the wave period is a deterministic variable and, therefore, the random variability of the wave power estimate is caused only by the variation in the ocean wave heights.

IV.6.4 Semi-empirical probability distribution

As shown in Eq. (4.45) and (4.46) for wave power respectively in deep and shallow water limits, wave power can be represented as a non-linear random variable having a quadratic relationship with a linear random variable, i.e. normalized wave height. The quadratic relationship is analogous to the mathematical model used to develop the semi-empirical model Eq. (2.3) with zero linear term $\alpha = 0$ and zero shifting term $\gamma = 0$. The semi-empirical approach being utilized here to model the probability distribution of wave power is a Rayleigh-Stokes model with $R = 4.0$. This is consistent with the assumption that linear waves have narrow-banded energy spectrum. The probability distributions of the special RS semi-empirical model are given in Eq. (2.46).

A major contribution of the semi-empirical models to the field of ocean wave energy is perhaps their application in estimation of random variability of the absorbed or converted wave power in a Wave Energy Convertor (WEC) device. The power extracted from ocean waves by a WEC is a function not only of the wave conditions at the offshore site, but also depends on the hydrodynamic characteristics of the device, the power-take-off system, the control strategy, etc. Considering all these variables explicitly in the theoretical probability distribution of absorbed power would be quite complicated and would require detailed design information for a particular WEC. That type of detailed information is proprietary and usually unavailable. Having samples of

extracted wave power by a WEC during a given seastate, one can apply a semi-empirical probability distribution to estimate the random variability of the extracted wave power.

IV.6.5 Sample data sets

The wave power sample data set for a sea-state can be obtained from directly analyzing time series developed from simulations or measured in the laboratory or the field. Basically, the observations are developed from a zero-crossing analysis which in turn allows a wave-by-wave estimate of the average wave power in each wave cycle [102]. More specifically, the wave periods T_i and associated wave heights H_i are obtained from the zero-crossing analysis of the surface elevation time series. Next, for each pair of H_i and T_i the wave power P_i is estimated using deepwater limit Eq. (4.43) or Eq. (4.44) for the shallow water limit, and consequently the wave power samples are developed. The samples are then normalized with the characteristic values $\hat{P}_d$ and $\hat{P}_s$ depending upon the limit selected. For illustrative purposes, the wave surface elevation time series were generated for the deep water limit using a JONSWAP wave amplitude spectrum model for different values of the peakedness parameter, specifically $\gamma_s = 1.0,\ 3.3,\ 7.0$, while the significant wave height $H_s = 4.0\,\mathrm{m}$, and peak period $T_p = 10.0\,\mathrm{s}$ were held constant. In order to reduce the sample size effects, for each sea-state 60 hours of surface wave elevation (about 26,000 waves) were generated with sample rate of 5 Hz employing uniformly distributed random phase in the range of $\left[0, 2\pi\right]$.

The JONSWAP model parameters, the spectral width, estimates of average and energy periods $T_{-1} = m_{-1}/m_0$, the ratio $\overline{T}/T_{-1}$, and estimates of the mean deepwater power $\overline{P}_d = \frac{\rho g^2}{64\pi} T_{-1} H_s^2$ and the ratio $\overline{P}_d/\hat{P}_d$ are presented in Table 30. The ratio $\overline{T}/T_{-1}$ depends on the wave spectral shape and width and is expected to converge to a value of unity for an ideal narrow-banded process. As shown in Table 30, the ratio of $\overline{T}/T_{-1}$ varies in the range of 0.90-0.94 for the sea-states specified. The mean wave power is routinely used as the characteristic of the seastate and is applied in conjunction with the seastate histogram to estimate the long-term probability distribution of wave power at a site. The values given for mean wave power in Table 30 indicate that waves with narrower energy spectra have higher mean power as the energy is focused around the peak period and the small wave heights are less probable in a narrow-banded process as compared to a process with wider spectrum. These results can be contrasted with the single value of characteristic wave power $\hat{P}_d = 78.89$ KW/m, obtained for each sea-state which yields an even higher estimate. The value of $\hat{P}_d$ represents the mean wave power in an ideal narrow-banded spectrum with zero width.

Table 30 Characterization of the simulated sea-states and the associated wave power resource.

Seastate No.	H_s (m)	T_p (sec)	γ_s	ν_s	T_{-1} (sec)	$\overline{T}$ (sec)	$\overline{T}/T_{-1}$	$\overline{P}_d$ (KW/m)	$\overline{P}_d / \hat{P}_d$
S-4	4.0	10.0	1.0	0.42	8.57	7.73	0.90	67.62	0.86
S-5	4.0	10.0	3.3	0.38	9.03	8.35	0.92	71.26	0.90
S-6	4.0	10.0	7.0	0.35	9.31	8.78	0.94	73.46	0.93

IV.6.6 Analysis and Results

The semi-empirical model used here takes advantage of two model parameters, i.e. β and γ, and therefore requires the sample estimates of the first two moments or L-moments for parameter estimation purposes. In Table 31, the first two moments and L-moments of the wave power samples are presented. The sample names in this table indicate the seastate the observations are generated from, e.g. the wave power sample S4WP contains the wave power observations of the simulated wave record from the seastate S-4. In this table, the statistics of the theoretical distribution, Eq. (4.50), and the narrow-band approximation, Eq. (4.52) are presented as well. As shown in Table 31, the first two moments (L-moments) of the wave power samples increase as the spectrum width decreases from sample S4WP to S6WP. A reasonable agreement between the normalized mean wave power values estimated from frequency domain analysis (see Table 30) and the mean wave power values obtained from wave-by-wave analysis in time domain (see Table 31) is observed. As shown in Table 31, the theoretical model tends to overestimate the sample statistics while its statistics get closer to the sample

statistics for wave records with narrower spectra. The narrow-band approximation constantly overestimates the sample mean and variance.

Table 31 L-moments and moments of wave power samples.

Statistics	Narrow-banded	S4WP		S5WP		S6WP	
		Sample	Theo.	Sample	Theo.	Sample	Theo.
$l_1(p_d)$	1.00	0.82	0.91	0.89	0.96	0.94	0.98
$l_2(p_d)$	0.50	0.42	0.59	0.45	0.55	0.46	0.53
$\hat{\mu}_1(p_d)$	1.00	0.82	0.91	0.89	0.96	0.94	0.98
$\hat{\mu}_2(p_d)$	1.00	0.66	0.89	0.75	0.94	0.80	0.95

Utilizing the sample statistics given in Table 31, the semi-empirical model parameters are estimated with application of MoL and MoM parameter estimation methods and the results are presented in Table 32. The estimates of parameter β from both parameter estimation methods are relatively close while MoL constantly estimates smaller shifting parameter γ. In Figure 52, Figure 53, and Figure 54, the PDF and CDF estimates of the semi-empirical models, theoretical model, and narrow-banded approximation are compared with the sample probability distributions of wave power samples S4WP, S5WP, and S6WP, respectively. As shown in these figures, the theoretical model is fairly accurate in representing the probability distribution of data and the model results are more accurate for wave records with narrower spectrum. The narrow-banded distribution constantly has a heavier tail than the sample distribution and the approximation overestimates the wave power statistics. The narrow-band approximation seems to be fairly accurate for waves with spectrum width $\nu_s < 0.4$. As

shown in Figure 52 to Figure 54, the semi-empirical distributions estimated by MoL and MoM are almost identical. More importantly, in the examples studied here, the semi-empirical model was found to be robust in capturing the probability distribution of data sets. The only deficiency of the semi-empirical model has been observed in estimation of small wave power ($p_d < 0.5$) probability distribution where the models overestimate the sample distribution PDF.

Table 32 Empirically estimated parameters of the semi-empirical model for the wave power samples.

Parameter	Method	S4WP	S5WP	S6WP
$\hat{\beta}$	MoL	0.105	0.111	0.115
	MoM	0.102	0.108	0.112
$\hat{\gamma}$	MoL	-0.013	0.002	0.014
	MoM	0.010	0.029	0.039

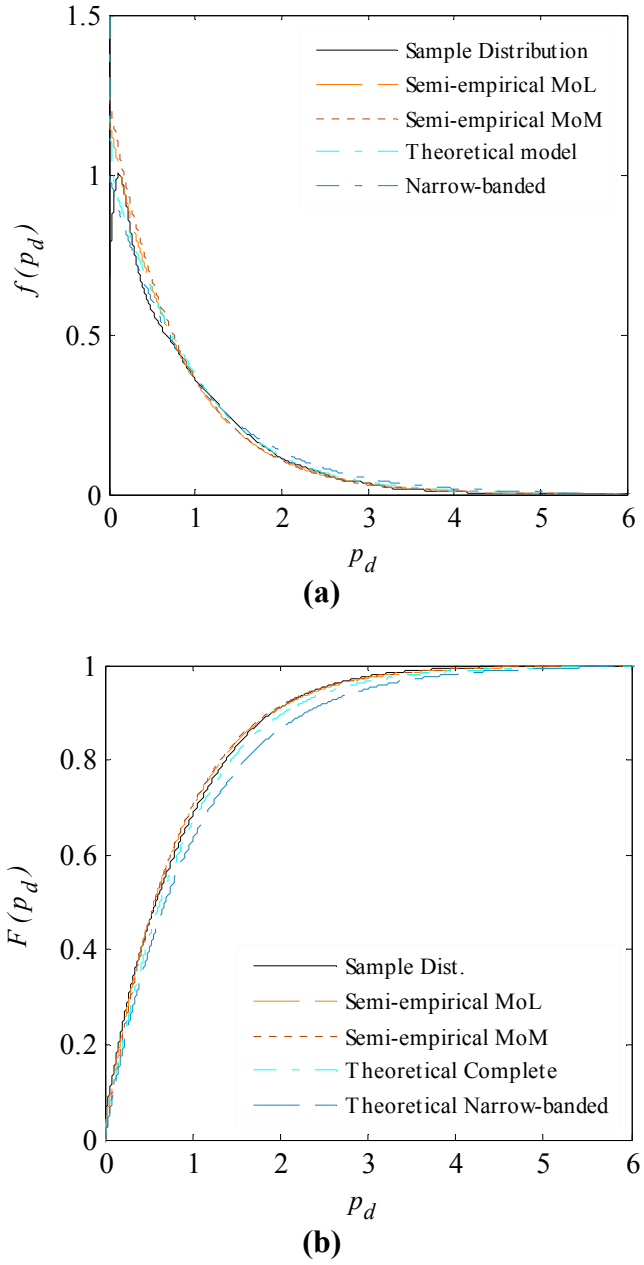
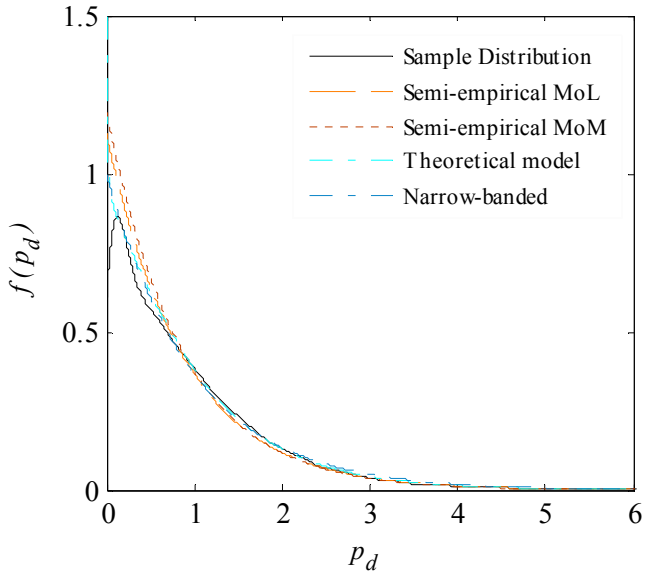
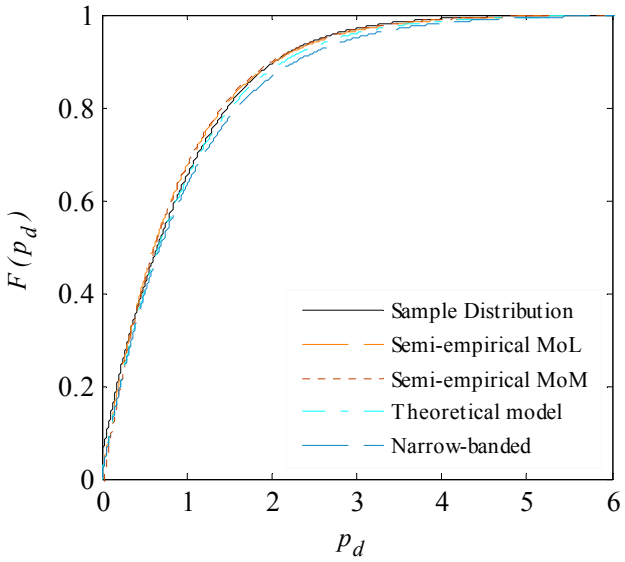


Figure 52 Probability distributions of wave power sample S4WP, (a) PDF, (b) CDF.

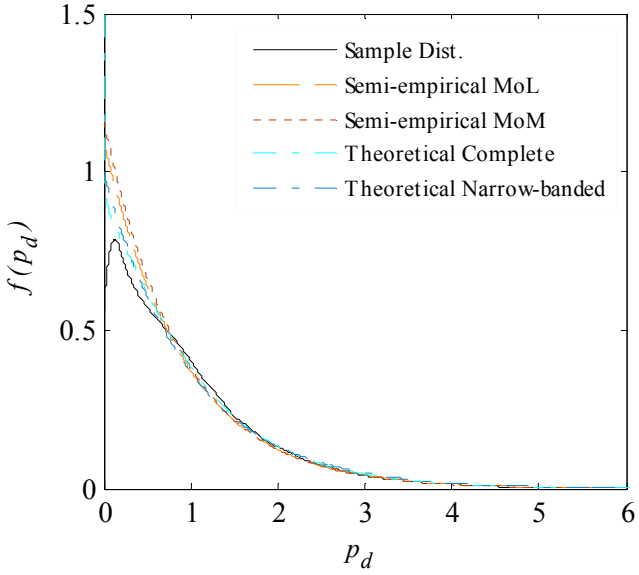


(a)

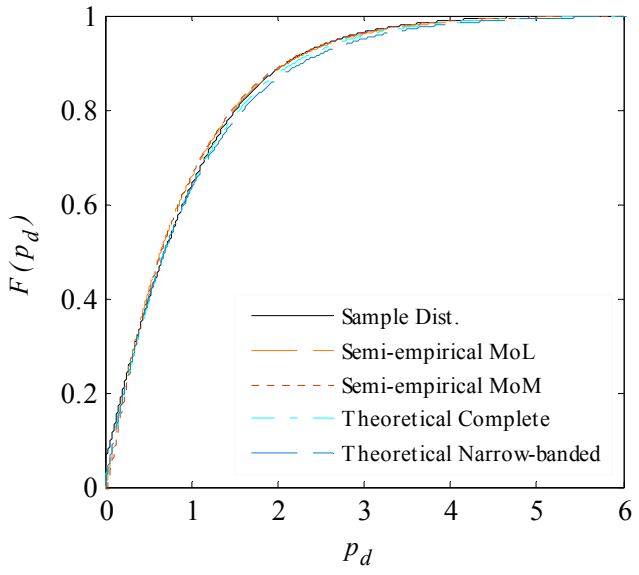


(b)

Figure 53 Probability distributions of wave power sample S5WP, (a) PDF, (b) CDF.



(a)



(b)

Figure 54 Probability distributions of wave power sample S6WP, (a) PDF, (b) CDF.

CHAPTER V

SUMMARY AND CONCLUSIONS

This study was focused on estimating the probability distribution of random variables in the field of wave-structure interaction. For this purpose, semi-empirical probability distribution models were introduced as alternatives to theoretical and empirical distribution models. In the semi-empirical probability distribution estimation approach, similar to the theoretical approach, the structural form of the parametric distribution model is developed based on a mathematical representation of the process. Similar to the empirical approach, the distribution model parameters of a semi-empirical model are estimated empirically by relating the model parameters directly to the sample statistics. As compared to the theoretical models, semi-empirical models have more connectivity to the sample data and are more flexible in capturing the probability distribution of data. In this matter, the semi-empirical approach behaves similarly to the empirical approach. In contrast with the complete empirical probability distribution, the structural form and parameters of the semi-empirical distribution model reflect some physical insight to the random variable. Considering these, the semi-empirical approach is the desirable approach for estimation of the probability distribution of complex non-linear random variables in different engineering fields.

In this study, three families of probability distributions, i.e. Gaussian-Stokes, Rayleigh-Stokes, and Weibull-Stokes, are developed for non-linear random variables. In development of the models, it is assumed that the non-linear random variable has a

quadratic relation with the corresponding linear variable. The quadratic transformation originated from the second-order perturbation expansion commonly known as Stokes expansion in the theory of wave mechanics. The second-order approximation considers the contributions of first-order, second-order and constant shifting with cooperation of the parameters α, β, and γ, respectively. The perturbation approximation has been routinely used in the field of offshore engineering to obtain an approximate solution for the weakly non-linear variables.

The quadratic transfer function is applied in conjunction with the random variable transformation rule to obtain the relation between the probability distributions of the linear and non-linear variables. Knowing that the linear process follows a standard Gaussian distribution law, the three-parameter Gaussian-Stokes distribution model was developed for weakly non-linear process. Similarly, the three parameter Rayleigh-Stokes model was developed for modeling the crests, troughs, and heights of a weakly non-linear process assuming that the linear random variable has a Rayleigh probability distribution. A more general form of the Rayleigh-Stokes model is developed assuming that the crest, troughs, and heights of the linear process follow a Weibull distribution law. The Weibull-Stokes model is a four-parameter probability distribution and has an additional unknown parameter, i.e. shape parameter, which is assumed to be a known constant in the Rayleigh-Stokes model. Furthermore, the application of the Rayleigh-Stokes and Weibull-Stokes models for extreme analysis is studied. More specifically, the statistics of the crest maxima are evaluated utilizing the semi-empirical probability distribution models. It is shown that that the asymptotic form of the crest maxima

distribution for large number of waves belongs to the Gumbel domain of attraction and the relation for the parameters of the asymptotic extreme distribution are presented in terms of the semi-empirical model parameters and the number of waves.

In general, the theoretical and semi-empirical models have a similar structural form, while the parameter estimation procedure differentiates the results of these models. In the theoretical approach, the model parameters are obtained based on a theoretical approximation of the process, e.g. second-order Stokes wave theory, and some simplifying assumptions, e.g. narrow-banded spectrum. On the contrary, the model parameters in the semi-empirical approach are estimated empirically from the sample data sets. In this study, two empirical parameter estimation methods, i.e. conventional method of moments and method of L-moments are utilized. In both methods, the parameter estimates are obtained from equating the distribution statistics with the corresponding sample estimators and solving the system of equations simultaneously for the unknown model parameters. In the conventional theory of statistics, moments are routinely utilized to characterize a probability distribution. More recently, distribution L-moments, developed from a certain linear combination of the probability-weighted moments, are utilized for a similar purpose. The main difference between the moments and L-moments is that the conventional moments give more weight to the tail part of the distribution. Therefore, the distribution moments are a better representative of the distribution tail as compared to the distribution L-moments. However, the sample L-moments are less biased than the corresponding sample moments and are less sensitive

to the contaminated observations. Additionally, the sample L-moments are the more efficient sample estimators for small samples.

As the first step of the empirical parameter estimation, the analytical relations between the model parameters and the distribution statistics, i.e. moments and L-moments, are derived for the three semi-empirical distribution models, i.e. Gaussian-Stokes, Rayleigh-Stokes, and Weibull-Stokes models. In case of the Gaussian-Stokes model, since the closed from solution of the model quantile is not available, the approximate distribution L-moments are obtained from application of Monte-Carlo simulations. For the semi-empirical models being studied here, the application of the method of L-moments is more straightforward as compared to the method moments. This is mainly because the relations between the distribution L-moments and the model parameters are considerably simpler. In case of method of moments, one requires an iterative solver scheme to obtain the parameter estimates of the three semi-empirical models.

In the third chapter, the sensitivity of the semi-empirical model statistics to the uncertainty of the model parameters is evaluated. It has been observed that the extreme statistics are highly sensitive to the variability of model parameters. Particularly, the variability of the parameter β is highly magnified on the distribution tail. This justifies the application of more robust parameter estimation method for extreme analysis. Moreover, the effect of sample variability on the estimates of the semi-empirical models is evaluated, and the performance of the method of moments and method of L-moments on samples with different sizes is compared. For this purpose numerous samples are

generated utilizing Monte-Carlo simulation technique. As a general conclusion, the sample L-moments were found to be less biased and have smaller variance as compared to the corresponding sample moments. It is confirmed that method of L-moments is the more efficient option for parameter estimation of small samples. In case of large samples, the uncertainty of the sample moments and sample L-moments is comparable. Although the Weibull-Stokes model is more flexible than Rayleigh-Stokes model in capturing the probability distribution of data, the uncertainty of the Weibull-Stokes model parameters is consistently higher. This is mainly because of the application of the fourth moment (fourth L-moment) in the parameter estimation of the Weibull-Stokes model and the more complex structure of this distribution model. Additionally, the Weibull-Stokes model statistics were found to be more sensitive to the sample size. It is concluded that the extreme statistics of the Rayleigh-Stokes and Weibull-Stokes models estimated from small samples are highly uncertain and need to be used with care. Moreover, illustrative examples are utilized to study the sensitivity of parameter estimation methods to the contaminated observations on the distribution tail. It is shown that sample L-moments are less affected by false extreme measurements and therefore the statistics obtained from this method are more reliable when the data quality is questionable.

In the final chapter, the semi-empirical models are applied to estimate the probability distribution of four non-linear random variables in the area of wave-structure interaction problem, i.e. wave surface elevation, wave crests, wave run-up over vertical columns, and wave power. The examples address analysis and design issues in both

extreme and benign environmental conditions. The semi-empirical models are evaluated over experimentally measured samples, as well as, numerically generated samples. The experimental data sets being used here are from a research study investigating the response of a mini-TLP structure in random seas.

In the first example, the Gaussian-Stokes model is used to estimate the probability distribution of the surface elevations of undisturbed ocean waves, as well as the surface elevations of disturbed waves in the vicinity and underneath an offshore structure. The analysis has been carried out for seastates with different spectral characteristics. The statistics of the Gaussian-Stokes model for undisturbed ocean wave samples are compared with the ones obtained from commonly used Gaussian and beta empirical model distributions and the well-known theoretical models. In case of disturbed wave elevation samples, the performance of Gaussian-Stokes model is evaluated quantitatively and compared with the performance of the empirical models. It has been observed that the Gaussian-Stokes model is robust in capturing the probability distribution of undisturbed wave data and its statistics compare well with the ones obtained from theoretical models. The Gaussian-Stokes model is reasonably accurate in capturing the general probability distribution of disturbed wave samples while it slightly overestimates the extreme statistics. The overestimation is caused by the fact that the perturbation expansion used in development of the semi-empirical model does not consider the effects of either the negative terms in the complete second-order expansion, or any energy loss mechanisms.

Rayleigh-Stokes and Weibull-Stokes semi-empirical models with the parameters estimated from method of moments and method of L-moments are applied to estimate the probability distribution of wave crests. In this example, the probability distributions of wave crests of undisturbed waves far from any obstacle, as well as, wave crests in the area close to the mini-TLP are evaluated. In case of the undisturbed crest samples, the statistics of the semi-empirical models are compared to the statistics of well-known theoretical models. Further, the semi-empirical model performance is compared to the Rayleigh and two-parameter Weibull empirical distributions. It has been observed that the results of the Rayleigh-Stokes and Weibull-Stokes semi-empirical models are fairly close and both methods are robust in representing the probability distribution of undisturbed crest samples. For these samples, the semi-empirical models constantly perform better than the theoretical and empirical models. The semi-empirical models found to be fairly accurate in representing the probability distribution of disturbed wave crests. The results of our analysis indicate that the method of moments has a slight advantage over the method of L-moments for extreme analysis when the sample is large enough and has reliable quality. In some studied samples, the tail distribution shows a slope change resulting in a flatter tail distribution. The Weibull-Stokes distribution model, because of the additional shape parameter, is more successful in capturing the tail distribution. The extreme statistics obtained from semi-empirical models are in a reasonable agreement with the extreme statistics obtained from the more advanced semi-parametric distribution.

In the third example, the Rayleigh-Stokes and Weibull-Stokes semi-empirical models are applied to estimate the probability distribution of wave run-up over vertical columns of the mini-TLP. The simplified theoretical models of linear and non-linear wave run-up found to be unsuccessful in representing the basic statistics of the wave run-up sample and consequently the theoretical model distributions are considerably different from the sample distributions. The empirical Rayleigh and two-parameter Weibull distributions tend to underestimate and overestimate the extreme statistics, respectively, and should be applied carefully. The semi-empirical models, however, are quite successful in capturing the complex probability distribution of wave run-up samples. The semi-empirical models perform better on weakly non-linear samples measured during benign seastates while their performance slightly worsens in case of the extremely energetic seastate. Similar to what we had for disturbed wave crests, the extreme wave run-up observations seem to have a different distribution than the small and medium wave run-ups. The semi-empirical models, similar to other parametric models, are more accurate in capturing the overall probability distribution of data and are less sensitive to the local changes in the probability distribution. This causes some differences between the sample tail distribution and the semi-empirical model tail. In general, the Weibull-Stokes model is more flexible in capturing the sample tail distribution when compared to the Rayleigh-Stokes distribution model, and consequently provides more realistic extreme statistics.

In the final example, a special form of the semi-empirical Rayleigh-Stokes model with no linear term is applied to model the probability distribution of wave power in

benign seastates. The semi-empirical model is evaluated for wave power samples obtained from wave-by-wave analysis of numerically generated wave samples. For this purpose large random linear wave realizations are generated from seastates with varying severity. The semi-empirical model statistics are compared with the ones obtained from complete theoretical distribution, as well as, the statistics obtained from narrow-banded approximation. It has been observed that the simple semi-empirical model is successful in capturing the probability distribution of wave sample and better represents the sample statistics when compared to the theoretical models.

REFERENCES

[1] Longuet-Higgins MS. On the statistical distribution of the heights of sea waves. Journal of Marine Research 1952; 11 (3): 245-66.

[2] Tayfun MA. Distribution of crest-to-trough wave heights. ASCE Journal of Waterway, Port, Coastal, and Ocean Engineering 1981; 107 (WW3): 149-58.

[3] Tayfun MA. Distribution of large wave heights. Journal of Waterway, Port, Coastal, and Ocean Engineering 1990; 116 (6): 686-707.

[4] Boccotti P. On mechanics of irregular gravity waves. Atti Accademia Nazionale dei Lincei, Memorie VIII 1989; 19: 111-70.

[5] Naess A. On the distribution of crest to trough wave heights. Journal of Ocean Engineering 1985; 12 (3): 221-34.

[6] Cartwright DE, Longuet-Higgins MS. The statistical distribution of the maxima of a random function. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 1956; 237 (1209): 212-32.

[7] Longuet-Higgins MS. The effect of nonlinearities on statistical distributions in the theory of sea waves. Journal of Fluid Mechanics 1963; 9: 193-217.

[8] Tayfun MA. Narrow-band nonlinear sea waves. Journal of Geophysical Research 1980; 85 (C3): 1548-52.

[9] Huang EN, Long SR, Tung CC, Yuan L, Bliven F. A non-Gaussian model for surface elevation of nonlinear random wave fields. Journal of Geophysical Research 1983; 88: 7597-606.

[10] Langley RS. A statistical analysis of nonlinear random waves. Journal of Ocean Engineering 1987; 14 (5): 389-407.

[11] Mori N, Yasuda T. A weakly non-Gaussian model of wave height distribution for random wave train. Journal of Ocean Engineering 2002; 29: 1219-31.

[12] Tayfun MA. Distributions of envelope and phase in weakly nonlinear random waves. Journal of Engineering Mechanics 1994; 120 (4): 1009-25.

[13] Tayfun MA. Statistics of nonlinear wave crests and groups. Journal of Ocean Engineering 2006; 33: 1589-1622.

[14] Arhan M, Plaisted, RO. Non-Linear deformation of sea wave profiles in intermediate and shallow water. Oceanologica Acta 1981; 4 (2): 107-15.

[15] Tung CC, Huang NE. Peak and trough distribution of nonlinear waves. Journal of Ocean Engineering 1985; 12 (3): 201-9.

[16] Fedele F, Arena F. Weakly nonlinear statistics of high random waves. Journal of Physics of Fluids 2005; 17 (1): 1-10.

[17] Kriebel DL. Non-linear run-up of random waves on a large cylinder. In: Proc. 12th international conference of ocean, offshore and arctic engineering, vol. 1. 1993. p. 49-55.

[18] Forristall GZ. On the statistical distribution of wave heights in a storm. Journal of Geophysical Research 1978; 83 (C5): 2353-58.

[19] Forristall GZ. Wave crest distribution: Observations and second-order theory. Journal of Physical Oceanography 2000; 30 (8): 1931-43.

[20] Stansell P. Distributions of freak wave heights measured in the North Sea. Journal of Applied Ocean Research 2004; 26: 35-48.

[21] Stansell P. Distributions of extreme wave crest and trough heights measured in the North Sea. Journal of Ocean Engineering 2005; 32: 1015-36.

[22] Niedzwecki JM, Van de Lindt JW, Gage JH, Teigen PS. Design estimates of surface wave interaction with compliant deepwater platforms. Journal of Ocean Engineering 2000; 27: 867–88.

[23] Indrebo AK, Niedzwecki, JM. Wave run-up on cylinders subject to deep water random waves. In: Proc. 23rd international conference of ocean, offshore and arctic engineering. 2004. p. 385-92.

[24] Myrhaug D, Leira BJ, Holm H. Wave power statistics for seastates. In: Proc. 28th international conference on ocean, offshore and arctic engineering. 2009. p. 809-16.

[25] Myrhaug D, Leira BJ, Holm H. Wave power statistics for individual waves. Journal of Applied Ocean Research 2009; 31(4): 246-50.

[26] Peebles PZ. Probability, random variables, and random signal principles. McGraw Hill; 2001.

[27] Rice SO. Mathematical analysis of random noise-and appendixes. Technical Publications Monograph B-1589. Bell Telephone Labs Inc., New York. 1950.

[28] Castillo E, Hadi AS, Balakrishnan N, Sarabia JM. Extreme value and related models with applications in engineering and science. Wiley Series in Probability and Statistics. John Wiley & Sons Inc. 2005.

[29] Van Gelder PHAJM. Statistical estimation methods in hydrological engineering. In: Proc. analysis and stochastic modeling of extreme runoff in Euro-Asian rivers under conditions of climate change. 2003. p. 11-57.

[30] Hosking JRM. L-moments: Analysis and estimation of distributions using linear combinations. Journal of the Royal Statistical Society. Series B Methodological 1990; 52 (1): 105-24.

[31] Greenwood JA, Landwehr JM, Matalas NC, Wallis JR. Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form. Journal of Water Resources Research 1979; 15 (5): 1049-54.

[32] Hosking JRM, Wallis JR. Regional frequency analysis an approach based on L-moments. Cambridge University Press; 1997.

[33] Dupuis DJ. Parameter and quantile estimation for the generalized extreme value distribution: a second look. Environmetrics 1999; 10: 119-24.

[34] Delicado P, Goria MN. A small sample comparison of maximum likelihood, moments and L-moments methods for the asymmetric exponential power distribution. Journal of Computational Statistics and Data Analysis 2008; 52: 1661-73.

[35] Öztekin T. Comparison of parameter estimation methods for the three-parameter generalized Pareto distribution. Turk Journal of Agriculture and Forestry 2005; 29: 416-28.

[36] Kundu, D, Raqab, MZ. Generalized Rayleigh distribution: Different methods of estimations. Journal of Computational Statistics and Data Analysis 2005; 49: 187-200.

[37] Pandey MD, Van Gelder PHAJM, Vrijling JK. The estimation of extreme quantiles of wind velocity using L-moments in the peaks-over-threshold approach. Journal of Structural Safety 2001; 23: 179-92.

[38] Park HW, Sohn H. Parameter estimation of the generalized extreme value distribution for structural health monitoring. Journal of Probabilistic Engineering Mechanics 2006; 21: 366–76.

[39] Efron B. Bootstrap methods: Another look at the jackknife. The Annals of Statistics 1979; 7 (1): 1-26.

[40] Efron B, Tibshirani R. An introduction to the bootstrap. Chapman & Hall/CRC; 1993.

[41] Caers J, Maes MA. Identifying tails, bounds and end-points of random variables. Journal of Structural Safety 1998; 20: 1-23.

[42] Dean RG, Dalrymple RA. Water wave mechanics for engineers and scientists. Advanced Series on Ocean Engineering. World Scientific; 2000.

[43] Hosking JRM, Wallis JR, Wood EF. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics 1985; 27 (3): 251-61.

[44] Hosking, JRM, Wallis JR. Parameter and quantile estimation for the generalized Pareto distribution. Technometrics 1987; 29 (3): 339-49.

[45] Leadbetter MR, Lindgren G, Rootzen H. Extremes and related properties of random sequences and processes. Springer Series in Statistics. Springer-Verlag; 1983.

[46] Silverman BW. Density estimation for statistics and data analysis. Monographs on Statistics and Applied Probability. Chapman and Hall; 1986.

[47] Silverman BW, Young GA. The bootstrap: To smooth or not to smooth? Biometrika 1987; 74 (3): 469-79.

[48] Falk M, Reiss RD. Weak convergence of smoothed and non-smoothed bootstrap quantile estimates. The Annals of Probability 1989; 17 (1): 362-71.

[49] Hall P, DiCiccio TJ, Romano JP. On smoothing and the bootstrap. The Annals of Statistics 1989; 17 (2): 692-704.

[50] Teigen PS, Niedzwecki JM, Winterstein SR. Wave interaction effects for non-compliant TLP. In: Proc. 11th international offshore and polar engineering conference. 2001. p. 453-61.

[51] Niedzwecki JM, Liagre P-YJ, Roesset JM, Kim MH, Teigen P. An experimental research study of a Mini-TLP. In: Proc. 11th international offshore and polar engineering conference. 2001. p. 631-4.

[52] Huang NE, Long SR. An experimental study of the surface elevation probability distribution and statistics of wind-generated waves. Journal of Fluid Mechanics 1980; 101 (1): 179-200.

[53] Denissenko P, Lukaschuk S, Nazarenko S. Gravity wave turbulence in a laboratory flume. Physical Review Letters 2007; 99 (1): 1-10.

[54] Ochi MK, Ahn K. Probability distribution applicable to non-Gaussian random processes. Journal of Probabilistic Engineering Mechanics 1994; 9 (4): 255-64.

[55] Winterstein SR. Nonlinear vibration models for extreme and fatigue. Journal of Engineering Mechanics 1988; 114 (10): 1769-87.

[56] Winterstein SR, Ude TC, Marthinsen T. Volterra models of ocean structures: Extreme and fatigue reliability. Journal of Engineering Mechanics 1994; 120 (6): 1369-85.

[57] Jha AK, Winterstein SR. Nonlinear random ocean waves: Prediction and comparison with data. In: Proc. ETCE/OMAE joint conference. 2000. p. 1-12.

[58] Ochi MK. Ocean waves: the stochastic approach. Cambridge University press; 1998.

[59] Srokosz MA. Notes and correspondence: A new statistical distribution for the surface elevation of weakly nonlinear water waves. Journal of Physical Oceanography 1998; 28: 149-55.

[60] Longuet-Higgins MS. On the joint distribution of wave periods and amplitudes of sea waves. Journal of Geophysical Research 1975; 80 (18): 2688-94.

[61] Longuet-Higgins MS. On the joint distribution of wave periods and amplitudes in a random wave field. Proceedings of Royal Society London 1983; 389: 241-58.

[62] Kac M, Siegert AJF. On the theory of noise in radio receivers with square law detector. Journal of Applied Physics 1947; 18: 383-97.

[63] Tayfun MA. On narrow-band representation of ocean waves: 1. Theory. Journal of Geophysical Research 1986; 9 (C6): 7743-52.

[64] Srokosz MA, Longuet-Higgins MS. On the skewness of sea-surface elevation. Journal of Fluid Mechanics 1986; 164: 487-97.

[65] Wei Wang DD, Fu Sun CQ. Comments on the surface elevation distributions derived by Huang. Journal of Applied Ocean Research 2002; 24: 185-8.

[66] Kriebel DL, Dawson TH. Nonlinearity in wave crest statistics. In: Proc. 2nd international symposium on ocean wave measurement and analysis. 1993. p. 61-75.

[67] Kriebel DL, Dawson TH. Nonlinear effects on wave groups in random seas. Journal of Offshore Mechanics and Arctic Engineering 1991; 113: 142-7.

[68] Al-Humoud J, Tayfun MA, Askar H. Distribution of nonlinear wave crests. Journal of Ocean Engineering 2002; 29 (15): 1929-43.

[69] Prevosto M, Forristall GZ. Statistics of wave crests from models vs. measurements, Journal of Offshore Mechanics and Arctic Engineering 2004; 126: 43-50.

[70] Sweetman B, Winterstein SR, Meling TS, Birknes J. Airgap prediction: Use of second-order diffraction and multicolumn models. In: Proc. 11th international offshore and polar engineering conference. 2001. p. 390-97.

[71] Sweetman B, Winterstein SR, Meling TS. Airgap prediction from second-order diffraction and Stokes theory. International Journal of Offshore Mechanics and Polar Engineering 2002; 12(3): 184–9.

[72] Sweetman B, Winterstein SR. Non-Gaussian air gap response models for floating structures. Journal of Engineering Mechanics 2003; 129 (3): 302–9.

[73] Izadparast AH, Niedzwecki JM. Estimating wave crest distributions using the method of L-moments. Journal of Applied Ocean Research 2009; 31: 37-43.

[74] Krogstad HE, Barstow SF. Analysis and applications of second-order models for maximum crest height. Journal of Offshore Mechanics and Arctic Engineering 2004; 126: 66-71.

[75] Nerzic R, Prevosto MA. Weibull-stokes model for the distribution of maximum wave and crest heights. In: Proc. 7th international offshore and polar engineering conference vol. 3. 1997. p. 367-77.

[76] Dawson TH. Maximum wave crests in heavy seas. Journal of Offshore Mechanics and Arctic Engineering 2000; 122: 222-4.

[77] Prevosto M, Krogstad HE, Robin A. Probability distributions for maximum wave and crest heights. Journal of Coastal Engineering 2000; 40: 329-60.

[78] Forristall GZ. Maximum wave heights over an area and the air gap problem. In: Proc. 25th international conference on ocean, offshore and arctic engineering. 2006. p. 11-15.

[79] Longuet-Higgins MS. On the distribution of the heights of sea waves: Some effects of nonlinearity and finite band width. Journal of Geophysical Research 1980; 85 (C3): 1519-23.

[80] MacCamy RC, Fuchs RA. Wave forces on a pile: Diffraction theory. Tech. Memo. 69. US Army Corps of Engineers 1954.

[81] Lighthill MJ. Waves and hydrodynamic loading. In: Proc. international conference on the behavior of offshore structures vol. 1. 1979. p. 1-40.

[82] Molin B. Second order diffraction loads upon three dimensional bodies. Journal of Applied Ocean Research 1979; 1: 197-202.

[83] Newman JN, Sclavounos PD. The computation of wave loads on large offshore structures. In: Proc. international conference on the behavior of offshore structures vol. 2. 1988. p. 605-22.

[84] Kim MH, Yue DK. The complete second-order diffraction solution for an axisymmetric body: Part 1. Monochromatic incident waves. Journal of Fluid Mechanics 1989; 200: 235-64.

[85] Newman JN. The second-order wave force on a vertical cylinder. Journal of Fluid Mechanics 1996; 320: 417-43.

[86] Lee CH, Newman JN. Computation of wave effects using the panel method. Numerical Models in Fluid-Structure Interaction 2004; 1: 1-41.

[87] Kriebel DL. Non-linear wave interaction with a vertical circular cylinder: Part I: Diffraction theory. Journal of Ocean Engineering 1990; 17 (4): 345-77.

[88] Kriebel DL. Non-linear wave interaction with a vertical circular cylinder: Part II: Wave run-up. Journal of Ocean Engineering 1992; 19 (1): 75-99.

[89] Niedzwecki JM, Duggal AS. Wave run-up and forces on cylinders in regular and random waves. Journal of Waterway, Port, Coastal, and Ocean Engineering 1992; 118 (6): 615-34.

[90] Mercier RS, Niedzwecki JM. Experimental measurement of second-order diffraction by a truncated vertical cylinder in monochromatic waves. In: Proc. international conference on the behavior of offshore structures vol. 2. 1994. p. 265-88.

[91] Nielsen FG. Comparative study on air-gap under floating platforms and run-up along platform columns. Journal of Marine Structures 2003; 16: 97-134.

[92] Kriebel DL. Non-linear wave run-up on large circular cylinders. In: Proc. of civil engineering in the ocean. 1992; 173-87.

[93] Izadparast AH, Niedzwecki JM. Probability distributions for wave run-up on offshore platform columns. In: Proc. 28th international conference of ocean, offshore and arctic engineering. 2009. p. 607-15.

[94] Izadparast AH, Niedzwecki JM. Probability distributions of wave run-up on a TLP model. Journal of Marine Structures 2010; 23 (2): 164-86.

[95] Sarpkaya T, Isaacson M. Mechanics of wave forces on offshore structures. Van Nostrand Reinhold Company Inc. 1981.

[96] Pontes MT. Assessing the European wave energy resource. Journal of Offshore Mechanics and Arctic Engineering 1998; 120: 226-31.

[97] Bedard R, Hagerman G, Previsic M, Siddiqui O, Thresher R, Ram B. Final summary report: Offshore wave power feasibility demonstration project. E2I EPRI WP 009-US Rev2. 2005.

[98] Henfridsson U, Neimane V, Strand K, Kapper R, Bernhoff H, Danielsson O, Leijon M, Sundberg J, Thorburn K, Ericsson E, Bergman K. Wave energy potential in the Baltic Sea and the Danish part of the North Sea, with reflections on the Skagerrak. Journal of Renewable Energy 2007; 32 (12): 2069-84.

[99] Defne Z, Haas KA, Fritz HM. Wave power potential along the Atlantic coast of the southeastern USA. Journal of Renewable Energy 2009; 34(10): 2197-205.

[100] Iglesias G, López M, Carballo R, Castro A, Fraguela JA, Frigaard P. Wave energy potential in Galicia (NW Spain). Journal of Renewable Energy 2009; 34(11): 2323-33.

[101] Izadparast AH, Niedzwecki JM. Estimating the potential of wave power resources. Journal of Ocean Engineering 2010 (accepted for publication).

[102] Smith GH, Venugopal V. A generic method for determining WEC power conversion from a random sea. In: Proc. international offshore and polar engineering conference. 2006. p. 460-5.

[103] Wand MP, Jones MC. Kernel smoothing: Monographs on statistics and applied probability 60. Chapman and Hall/CRC; 1994.

[104] Pickands III J. Statistical inference using extreme order statistics. The Annals of Statistics 1975; 3 (1): 119-31.

APPENDIX A

SEMI-PARAMETRIC BOOTSTRAP ANALYSIS

In the standard non-parametric bootstrap method the sample probability distribution $\hat{X}$ is used as the distribution model to generate independent samples with replacement [39, 40]. A bootstrap sample of size n is drawn with replacement as

$$X_j^* = \hat{X}\left(u_j\right) \qquad j = 1, 2, ..., n \tag{A.1}$$

where, $\hat{X}$ is the sample quantile function and u is a random number distributed uniformly in the range of $(0-1)$. The bootstrap samples are used to obtain bootstrap estimates ϕ_m^* of a sample estimate ϕ_s of a parameter of interest ϕ. The set of estimates $\phi^* = \left[\phi_1^*, \phi_2^*, ..., \phi_M^*\right]$, where M is the number of bootstrap sample, constitutes the sampling distribution of ϕ_s. It can be shown that the bootstrap estimate of true bias $\left(E\left(\phi_s\right) - \phi\right)$ is

$$\Delta_{\phi_s} = \frac{1}{M}\sum_{m=1}^{M}\phi_m^* - \phi_s = \left(\overline{\phi}^* - \phi_s\right) \tag{A.2}$$

Similarly, the bootstrap estimate of true variance and root-mean-square error (RMSE) of ϕ_s are:

$$\sigma_{\phi_s}^2 = \frac{1}{M-1}\sum_{m=1}^{M}\left(\overline{\phi}^* - \phi_m^*\right)^2 \tag{A.3}$$

$$RMSE\left(\phi_s\right) = \sqrt{\Delta_{\phi_s}{}^2 + \sigma_{\phi_s}^2} \tag{A.4}$$

The empirical distribution $\hat{X}$ used in the standard bootstrap method is a discrete distribution, in that, samples will be drawn from the original sample values and consequently nearly every sample will contain repeated values. The smoothed bootstrap is a modification to the standard bootstrap procedure in which the bootstrap samples are generated from a smoothed version of $\hat{X}$ [47]. The smoothed distribution is commonly estimated utilizing the kernel density estimation that is the most appealing non-parametric method for probability distribution estimation. The standard kernel density $\hat{f}_k$ of a univariate density f of a random sample $X_1,\ldots,X_n$ of size n is [46]

$$\hat{f}_k(x)=(nh)^{-1}\sum K\left\{h^{-1}(x-X_i)\right\} \tag{A.5}$$

where, $K(.)$ is a kernel function and h is the bandwidth. Kernel function is a probability distribution which peaks at the estimation point x and decreases as it gets further from the estimation point. Standard normal distribution is a common option for this purpose. The bandwidth is a critical parameter in this method and here an estimate of the bandwidth is obtained using a well-known direct plug-in method [103].

To generate bootstrap samples for extreme quantile estimation the tail part of the kernel distribution is modified by fitting a Generalized Pareto Distribution (GPD) to the n-extreme values over a given threshold ς_p [41]. The Bootstrap sampling is performed from the semi-parametric probability distribution $\hat{F}_s$,

$$\hat{F}_s\left(x\middle|\varsigma_p\right)=\begin{cases}\left(1-\hat{F}_k\left(\varsigma_p\right)\right)\hat{F}_p(x)+\hat{F}_k\left(\varsigma_p\right) & x>\varsigma_p\\ \hat{F}_k\left(\varsigma_p\right) & x\le\varsigma_p\end{cases} \tag{A.6}$$

where, $\hat{F}_p$ is the GPD fitted to the peaks over the threshold ς_p, specifically, [104]

$$\hat{F}_p(x) = \begin{cases} 1 - \left(1 + \kappa_p \dfrac{x - \varsigma_p}{\delta_p}\right)^{-1/\kappa_p} & if \quad \kappa_p \neq 0 \\ 1 - \exp\left(\dfrac{x - \varsigma_p}{\delta_p}\right) & if \quad \kappa_p = 0 \end{cases} \tag{A.7}$$

for

$$\begin{cases} x \geq \varsigma_p & if\ \kappa_p \geq 0 \\ \varsigma_p \leq x \leq \varsigma_p - \delta_p / \kappa_p & if\ \kappa_p < 0 \end{cases}$$

The unknown parameters of the GPD distribution, including the scale parameter δ_p and the shape parameter κ_p, are estimated using the method of L-moments [32]. In order to obtain the optimum threshold ς_p the methodology recommended by Caers and Maes [41] is applied.

VITA

Amir Hossein Izadparast

CE/TTI 802-B, 3136 TAMU, College Station, TX 77843-3136
e-mail: amirizad@tamu.edu

Amir Hossein Izadparast received his B.S in civil engineering from Shiraz University, Shiraz, Iran in 2002. He got his M.S. in marine structures from University of Tehran, Tehran, Iran in 2005. Later on, he started his doctorate studies in ocean engineering in Civil Engineering department at Texas A&M University and graduated with his Ph.D. in December 2010.

Before starting his Ph.D. studies, he had worked as a professional civil structural and marine structural designer. During his Ph.D. studies, he had the chance to participate in a summer internship with British Maritime Scientific Marine Services, mainly working on advanced data analysis.

Amir's research interests are focused on broadening the range of statistical analysis methods applied to offshore engineering problems and the probabilistic analysis and design of offshore structures. Additionally, he is interested in data analysis of ocean engineering and physical oceanography data.